A Novel Approach to Mapping Flooding Extent in the Chobe River Basin from 2014 to 2016 Using A Training Library

by

Mitchell P. Braget

B.S., University of North Dakota, 2015

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF ARTS

Department of Geography
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Approved by:

Major Professor
Douglas G. Goodin

Copyright

© Mitchell P. Braget 2017.

Abstract

The Chobe River Basin (CRB) is a flood-dependent ecosystem that relies on seasonal floods from the Zambezi and Linyanti Rivers. These flood pulses provide water for the flood recession agriculture in the region, water for the fishing grounds around Lake Liambezi, and nutrients for the vegetation in the CRB. Recent years have shown an increase in the magnitude of flooding, which could have consequences on the region's biodiversity and the people living in the CRB. The goal of this study is to develop a classification framework based on a training library and time-windows to use in classifying the extent of flooding in the CRB. MODIS MOD09A1 satellite imagery served as the satellite imagery. Bands one through seven were converted into the tasseled cap transformation to serve as the feature selection. The study period, from February to July, is broken down into three time-windows. The time-windows are used because the land covers in the CRB go through significant spectral changes during the study period and the three time-windows seek to improve the classification accuracy. The classification methods include maximum likelihood classifier (MLC), decision trees (DT), and support vector machines (SVMs). The results show that DT and SVMs provide the highest overall accuracy and kappa values over MLC. Classification using the time-window method was statistically significant when comparing kappa values and visually, images classified using the correct training library for a time-window displayed higher agreement with the reference data. Flooding extent was high for 2014 but low in 2015 and 2016, indicating a decreasing trend. DTs provided better inundation maximums compared to SVMs and therefore is the reason that DT are the best classification technique. The results will provide planners with information regarding the extent of flooding in the CRB and where waterborne diseases occur in the region. A new classification technique is also developed for the remote sensing literature.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Dr. Douglas Goodin for spending copious hours editing this thesis and helping me write in a more technical manner. Your support through the writing process was invaluable, even when I wanted the writing to just be done already. Thanks to my committee members of Dr. Jida Wang and Dr. Shawn Hutchinson for suggestions and help with concepts I was unsure about. I also thank the Geography Department here at Kansas State University for giving me the change to pursue a Master's degree. Thank you to the graduate students within the Department of Geography. Your criticisms and shenanigans kept me sane throughout my time here.

I also thank my parents for giving me the opportunity to seek a Master's degree far from home, even though my mom was worried about Kansas = Tornadoes. To my older brother Mike, thanks for your advice about graduate school and how to cope with its difficulties. To my twin Austin, thank you for experiencing graduate school with me and providing an outlet for my frustrations.

This research was funded by the National Science Foundation Dyn1518486 grant. The grant allowed me to work on the project over the summer and provided financial security. Kathy Alexander of Virginia Tech University was the lead PI for the ongoing research in the Chobe River Basin. Thank you to Kathy and again to Douglas Goodin for allowing me to work on this project and contribute to research in Africa.

Chapter 1 - Introduction and Literature Review

Wetlands and seasonally inundated floodplains in semi-arid environments play a large part in the hydrologic cycle, as well as human diseases and livelihood support systems through controls such as fisheries productivity, water availability, and flood recession agriculture (Pricope, 2013). Wetlands also provide food, sediment storage, flood control, recreation, recharging of groundwater, erosion control, and wildlife habitat (Gallant, 2015; Halabisky et al., 2011). It is estimated that one billion people live in extreme flood areas, and that number may double by 2050 due to climate change and population increase (Long et al., 2014). With population increasing in urban areas, which can be next to wetlands and other hazardous areas, it is important to understand the dynamics of flooding extent to provide better information for those people located in high-risk flooding areas.

Since its inception in the mid-1970s, civil remote sensing has been used to identify large scale areas of potential risk. Remote sensing allows for large-scale area mapping and analysis without requiring large amounts of fieldwork. In addition, remotely sensed data can provide a temporal analysis of how an area changes over time (Burke et al., 2016). However, wetlands are difficult to monitor due to a number of factors. These factors include; presence of a large number of land-cover types or vegetation, large variety of reflectance and energy backscatter properties due to the mixture of water and vegetation, and steep environmental gradients in and around the edges of wetlands, which produce mixing of the spectral signatures (Gallant, 2015).

One of the difficulties of classifying remotely sensed imagery, is the need to collect training samples for each individual image. How do you approach the problem of classifying a series of images over a long period of time that requires a large amount of imagery? In this thesis, the flooding extent of the Chobe River Basin (CRB), a flood-pulse dependent basin located on the border between Botswana and Namibia, is mapped using a time-window framework based on a training library. I explored the temporal evolution of flooding extent in the CRB from 2014 to 2016 to determine the effectiveness of this classification framework. Previous studies have shown that the magnitude of flooding in the region is increasing (Burke et al., 2016); this study will quantify the spatial extent of flooding in the CRB to build upon previous work and determine if the flooding extent is increasing or decreasing. Remote sensing and other geospatial technologies have been effective tools for quantifying flooding extent

(Pricope et al., 2013; Li et al., 2015; Chen et al., 2013) and will function as the tool for analysis. Information gained using geospatial techniques will provide greater knowledge to policy makers managing the area around the CRB, as they evaluate the extent of flooding in the region.

This research was undertaken as part of a multi-investigator project to consider the role of human environmental disease dynamics in the Chobe River basin. The project integrates empirical studies of a dryland river system by modeling dynamical modeling of the environment and human system components of diarrheal disease for forecasting ability. Project findings will have application across fresh water and marine systems where floodplains exercise important influences on human well-being. Outputs from the project will have a broad application to a variety of critically important waterborne and water-associated pathogens, as for example, cryptosporidiosis, schistosomiasis, or malaria.

1.1 Historical Approaches to Flood and Wetland Mapping

There are a number of different techniques to quantify the extent of wetland inundation and flooding. These techniques include but are not limited to; thresholding images using histograms, use of different indices to enhance water features in images, and classification schemes that include supervised and unsupervised techniques. Some of these techniques are more efficient than others, however, and here a more thorough review is presented.

1.1.1 Water Extent Mapping Using Indices

A common first step in mapping floods and wetlands is to construct a remote sensing index. These indices are commonly called water indices and there are several varieties of them based on different combinations of multispectral bands. In other words, these indices take a multispectral band (e.g. red) and then divide it by another multispectral band, NIR. Another way to combine bands is to take red plus NIR divided by red minus NIR, which also happens to be the Normalized Difference Vegetation Index (NDVI). The indices go by a number of different names, such as the Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI), Modified Normalized Difference Water Index (MNDWI), and the Normalized Difference Pond Index (NDPI), just to name a few (Boschetti et al, 2014; Ogilvie et al., 2015). The difference between the indices come from the use of different multispectral band combinations to construct the indices. For example, the NDWI uses Green and Near-Infrared (NIR) bands to pick up open water features and eliminate the presence of soil and vegetation

features (McFeeters, 1996), while the MNDWI uses Green and Shortwave-Infrared (SWIR) to better differentiate open water from land noise (Xu, 2006). Different spectral bands are effective at picking up different reflectances of features. For example, Boschetti et al. (2004), found that indices that include NIR-SWIR (800 vs 1640/2130nm) were good at monitoring plant water content, while indices that include VIS-NIR (550/670 vs. 1240nm) were good at detecting water/non-water conditions.

The MNDWI is a very popular flooding index because of its ability to detect a wide range of water bodies (Ogilvie et al., 2015; Huang et al., 2014; Chen et al., 2013). Some of the drawbacks of using the MNDWI include, as floodwater recedes and vegetation increases, MNDWI tends to underestimate water surfaces or shallow water that has vegetation underneath (Ogilvie et al., 2015). Others have found that using SWIR (1230 – 1250nm) works best at producing low thresholds that represented water because SWIR was able to differentiate water from the soil and vegetation (Ji, Zhang, and Wylie, 2009). In essence, it is very difficult to create an index which is sensitive to a single phenomenon and insensitive to other factors (Govaerts et al., 1999).

Another way of detecting water in images involves synthesizing information contained in multiple spectral bands. The tasseled cap transformation introduced by Kauth and Thomas (1976) compresses spectral data into a few bands, biophysically defined which can then be associated with physical parameters of the surface. The three components of the tasseled cap are brightness, greenness, and wetness. In a paper by Li et al. (2015), the authors used the brightness component to delineate water and inundated pixels in wetlands. The brightness component is a weighted sum of all six bands for Landsat (seven for MODIS) and correlated to moisture content of soils (Li et al. 2015). The authors found that their Tasseled Cap Brightness Index (TCBI) was effective at monitoring wetness in wet and dry years and suggested using the TCBI in seasonally flooded wetlands in semi-arid areas for further research in its effectiveness.

1.1.2 Thresholding Techniques

Using an index to map water extent is usually just the first step in flood and wetland mapping because extra manipulation of the index is required to further highlight water in images. Thresholding techniques are another way of delineating water extent. The basic idea with using thresholds is that in a histogram of an ROI in an image, one part is water and the other is land. Figure 1.1 displays a frequency histogram of a Normalized Difference Water Index (NDWI)

Landsat 7 image that researchers would use to set a threshold. The red line represents an optimal threshold, where values to the right of the line represent water and values to the left represent land. It can be difficult to define an optimal value that delineates the two land covers. A number of studies have utilized thresholding techniques after selecting an index to classify water pixels.

Pricope (2013) made use of MODIS MOD13Q1, which consists of 16-day Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) images, and Advanced Very High Resolution Radiometer (AVHRR) imagery to determine if a reduction in flooding extent occurred from 1985 to 2010. Using the EVI image, Pricope used a threshold to delineate open water and flooded vegetation pixels. Pricope found that the thresholding picked up open water well but also overestimated due to wildfires. Long et al. (2014) used Synthetic Aperture Radar (SAR) in a study on the Chobe River Basin that used thresholding as part of their methods in discerning what pixels were water in each image. Then, using a decision tree classification, the threshold was applied along with a digital elevation model to further classify water efficiently. The authors found that SAR imagery was able to differentiate flooding in vegetated pixels due to the brightening effects by the SAR signal's double-reflectance off of objects in standing water (Long et al., 2014). Amarnath and Rajah (2015) used MODIS MOD09A1 imagery, which is 8-day composite imagery, to determine the flooding extent of the Indus River basin. They used a threshold method to identify the flooded pixels and validated their data with the advanced land observing system (ALOS) sensor. These authors found that their MODIS imagery included non-flooded areas due to the coarse resolution but the overall classification of inundated areas was a success.

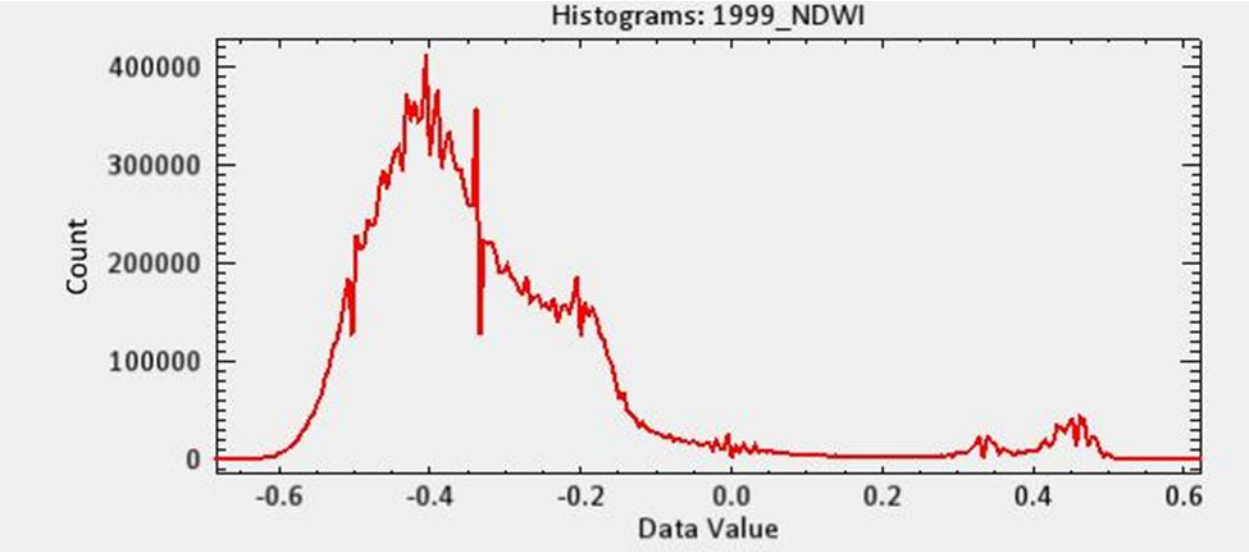


Figure 1.1 Frequency distribution of an NDWI Landsat 7 image

Threshold analysis also incorporates indices and ancillary GIS data to better pick up flooding. Chen et al. (2014) used MODIS 8-day composite images, as well as the MNDWI to help with their threshold analysis. The authors also used DEMs and flow data in their threshold analysis to better understand flood inundation extent. The results showed that the coarse resolution of MODIS limited estimates of inundation extent and the MNDWI underestimated the extent of small channels and patches of inundation (Chen et al., 2014). Rokni et al. (2014) calculated a number of indices (e.g. NDWI, MNDWI, etc.), then used image-specific thresholding to detect lake surface area. The authors used Principal Components Analysis (PCA) of NDWI images and applied thresholding to manually classify the water in the PCs with a single threshold over multiple dates. They found that their method efficiently detected the surface water change between two and three image dates while only having to define a single threshold, which saved time in the long run (Rokni et al., 2014).

1.1.3 Classification Techniques

One last way to delineate flooding extent is through classification techniques. Efficient supervised classification techniques address four challenges laid out in a paper by Millard and Richardson (2015). The first is to mitigate the Hughes phenomenon, or curse of dimensionality, which is where the number of variables is larger than the number of training samples. The second is to handle the nonlinearity of variables. Third, the classifier has to deal with imbalanced training samples and noise in the training and testing data. Finally, the classifier

must reduce the computation time. Classification techniques include both supervised and unsupervised classification.

Supervised classification techniques are where training samples are taken to train the classifier in classifying the image. One example of a supervised classification technique that has gained a wide use is random forest. Random forest (RF) is an ensemble supervised classification algorithm that uses randomly selected features to build a user-defined number of decision trees for classification of categorical and continuous data (Brieman, 2001). RF has only two user-defined parameters, number of trees to generate and the number of feature to split at each node, both of which are easy to set and is a reason the technique is increasing in usage. In comparison studies between RF and other classification techniques, RF scores highly in terms of accuracy and is comparable to support vector machines, which require more difficult user-defined parameters (Eisavi et al., 2015; Adam et al., 2014; Waske and Van der Linden, 2008; Li et al., 2013).

Unsupervised classification techniques do not require training samples to train the classifier and instead relies on the search for natural grouping, or clustering, of pixels. The majority of the work comes after the clustering, in which the analyst must assign spectral classes to information classes of interest. Unsupervised classification techniques include ISODATA and K-Means. ISODATA (Iterative Self-Organizing Data Analysis Technique) is an iterative process that takes several passes through the data set until there is little change in class assignment (Dechka et al., 2002). In other words, ISODATA can assist with finding spectrally homogenous clusters of pixels (Dechka et al., 2002). Therefore, ISODATA helps with an initial classification to understand the spectral signatures in a study area. ISODATA can also help with post-classification by looking at how well the initial classifier performed compared to the clustering of the ISODATA technique (Schmid et al., 2004). In essence, unsupervised classification techniques can assist with the classification process but should not be the only technique used. The major downfall of unsupervised classification techniques is that the clustering can at times be difficult to decipher what classes represent what on the ground.

1.2 Flood and Wetland Classification Techniques

I use the supervised classification techniques of maximum likelihood, support vector machines (SVMs), and decision trees (DTs) for classification of the water extent in the CRB.

Maximum likelihood and DTs have been widely used in the literature (Adam et al., 2014; Otukei and Blaschke, 2009; Srivastava et al., 2012) but SVMs have historically been difficult to implement. The maximum likelihood classifier is used because of its accessibility in a variety of software suites. DTs and SVMs are used because they are non-parametric techniques, which means they make no assumptions about the distribution of the training data for the classification, and because the training data is not completely normally distributed (Figure 1.2).

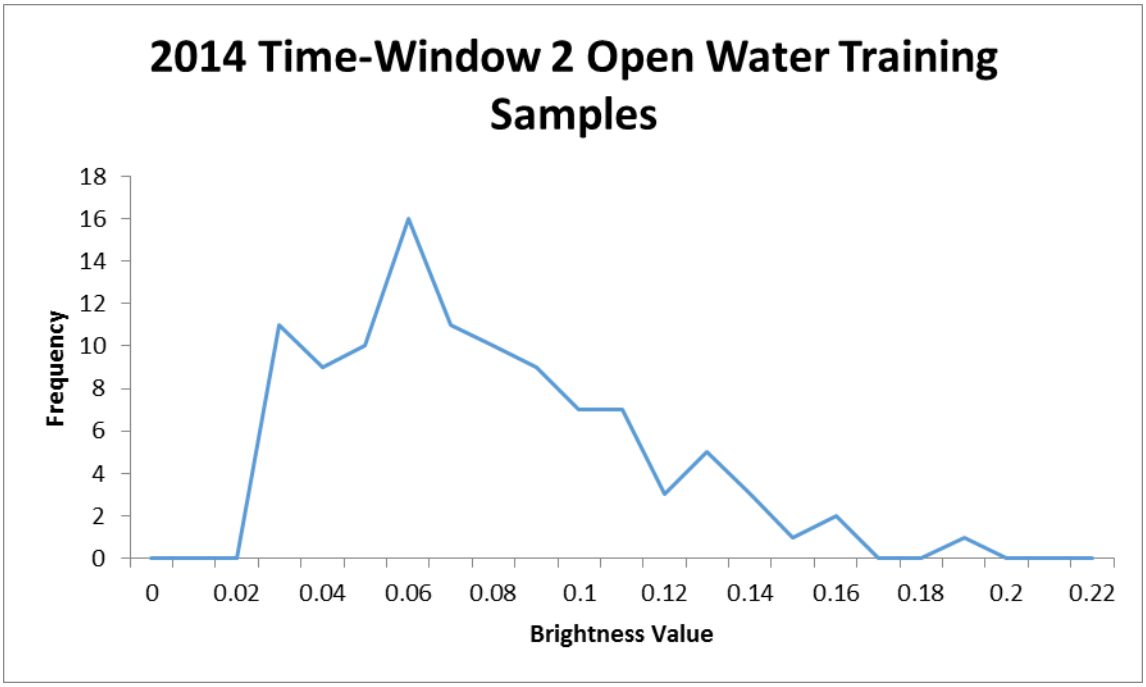


Figure 1.2 Open water training samples using the tasseled cap brightness band

1.2.1 Maximum Likelihood

Maximum Likelihood classifier (MLC) is a parametric supervised classification technique, which means it assumes the data is normally distributed. MLC uses the training data as the method of estimating mean vectors and variances of the classes, which work to estimate the probabilities through probability distribution functions (PDFs) (Srivastava et al., 2012). Once calculated, the PDFs assign a given pixel to a specific class of land cover based on its highest membership probability value (Pal and Mather, 2003).

MLC was chosen to classify imagery because it has a long history and therefore is in a large number of studies to serve as a benchmark against other and newer classification techniques (Pal and Mather, 2005; Hansen et al., 1996; Bischof et al., 1992; Liu et al., 2016). In one study, MLC was compared against artificial neural networks (ANNs) and SVMs, to

determine the best classification technique for land cover change over time (Srivastava et al., 2012). The results showed that ANNs provided the best overall accuracy with SVMs close behind and then MLC. In another study, MLC was used to compare classification accuracy against decision trees and SVMs in terms of land cover change (Otukei and Blaschke, 2010). The authors found that DTs performed better than MLC and SVMs. In a study by Shaker et al. (2012), panchromatic SPOT imagery was used to determine flooding extent and the results of the study concluded that the MLC classifier accurately modeled the flooding extent of the Nile River.

Advantages of working with MLC is that it is the most commonly used supervised classification technique, which allows for multiple comparison studies to look at, as well as easily accessed software to run the technique. Another advantage is that MLC considers not only the class average (mean) values in assigning a thematic category, but also the variability of brightness values within each class (Srivastava et al., 2012). MLC also looks at the probability of a pixel belonging to multiple classes. Limitations of MLC include that it is time and computationally intensive, as well as it requires the data to be normally distributed. One final point about MLC is that it requires a large number of training samples and they need to be pure or well-trained samples for an accurate classification (Otukei and Blachke, 2010).

1.2.2 Support Vector Machines

Support vector machines (SVMs) are a non-parametric, supervised statistical machine learning technique that has gained popularity in their ability to classify geospatial and remote sensing data. Non-parametric means that SVMs make no assumptions about the underlying data distribution, which makes SVMs particularly appealing to geospatial analysts, whose data is rarely normally distributed. SVMs aim to determine the optimal hyperplane in a set of data by using support vectors to maximize the margin or distance between the support vectors and the hyperplane (Waske and Van der Linden, 2008). The term maximum margin classifier can also describe SVMs due to the support vectors changing the distance of the hyperplane. For the most part, SVMs are linear classifiers that seek to classify two sets of data with the hyperplane (Cortes and Vapnik, 1995). Another type of hyperplane is the soft-margin hyperplane, which allows for some misclassification of the data (Cortes and Vapnik, 1995). There are techniques for cases that are not linearly separable, which rely on mapping the data into higher-dimensional feature

space and applying a kernel function to decrease the computational demand (Kavzoglu and Colkesen, 2009).

Research shows that SVMs have proven to be effective in classifying wetlands. In a study by Duro et al. (2012), SVMs along with Random Forest were found to provide a more adequate depiction of wetland areas. SVMs combined with 4 spectral bands and an NDVI layer were able to accurately portray the heterogeneity of different wetland areas in the Florida Everglades (Szantoi et al., 2013). Han et al. (2015), found that SVMs were able to define long-term changes for wetlands in Poyang Lake of China. Another study focusing on the Poyang Lake region found that SVMs effectively contributed to identifying the distribution of wetland vegetation plant functional types, using time series vegetation-water index images (Wang et al., 2012). SVMs were also used to identify community and species-level vegetation in wetlands with hyperspectral imagery (Zhang and Xie, 2013). The authors found that their SVM model was able to map vegetation in heterogeneous wetland ecosystems with an 85% overall accuracy rating.

SVMs have several advantages over other classification techniques. One of the major advantages of SVMs is their superior classification accuracy with small amounts of training data (Mountrakis et al., 2011; Pal, 2008; Adam et al., 2014). The high classification accuracy comes from the fact that SVMs make no assumptions on the distribution of the data, which reduces the error rate on the data that needs to be classified. Another advantage of SVMs is that they tend to not get trapped in some local minima because of the convexity of the cost function enables the classifier to always identify the optimal solution (Mountrakis et al., 2011). Simply put, SVMs always find the global minimum with quadratic computing. Finally, SVMs, unlike random forest, does not need to repeat the classifier training using the different random initialization of the training data, which can cut down on computation time (Mountrakis et al., 2011).

The main disadvantage with SVMs is that the selection of the two parameters C and γ can be difficult (Mountrakis et al., 2011). The problem of selecting the two parameters has recently been addressed with the LIBSVM library, which allows for a grid-search using V-fold cross-validation of these two parameters but can be difficult to use at times (Chih-Chung and Lin, 2011). SVMs also have a problem with noisy data or outliers in the data, which can result in a significant decrease in the performance of the SVM (Mountrakis et al., 2011). Therefore, it is imperative that quality training and testing sampling are made during the initial sampling stage

of a project. One final disadvantage has to do with the high computational costs of SVMs as the dimensionality of data increases (Mountrakis et al., 2011). These high computation costs have been addressed with the kernel functions in recent years but SVMs can still have high computational demands with larger study areas.

1.2.3 Decision Tree

Decision Trees (DTs) are a non-parametric supervised classification technique that does not make assumptions about the distribution of data. Decision trees work by having one root node, several internal nodes, and finally the terminal nodes, which are the final land covers (or other class property) that the pixels are assigned to (Otukei and Blaschke, 2010). Basically, the data is moved down the decision tree based on a user-defined classification framework. A pixel value will start at the root node and will go left or right to another node based on if the value is greater or smaller than the specified value. The process continues until that pixel value reaches a terminal node and it becomes part of a land cover class (i.e. Wetland).

DTs can be univariate or multivariate. A univariate DT is where the decision boundaries at each node in a tree are defined by the result of the test applied to a single feature and is evaluated at each internal node (Pal and Mather, 2003). A multivariate DT uses a linear combination of features when it splits the data into nodes (Pal and Mather, 2003). In other words, multivariate splits incorporate two variables at each node to determine the best split, whereas univariate only uses one variable to split at the node. One study found that multivariate DTs are more compact and have higher classification accuracy (Pal and Mather, 2003). I use a univariate approach with the three Kauth-Thomas transformation bands of brightness, greenness, and wetness as the decision on which to split each node. The reason for using a univariate instead of a multivariate DT is because while I have three variables, which are the brightness, greenness, and wetness bands from the KT transformation, only one variable is used at each node to partition the data.

DTs have seen use in a number of studies and researchers have found that they produce high accuracies when comparing them to minimum distance to means classifiers (Otukei and Blaschke, 2010). Another reason DTs continue to have use in the remote sensing community include their insensitivity to noisy input data (Funkenberg et al., 2014). In terms of recent studies, one found that DTs have comparative accuracy to random forest, SVMs, and object-based image analysis, although they had more omission and commission errors (Duro et al.,

2012). Wright and Gallant (2007) found that DTs combined with ancillary environmental data produced higher accuracy of wetland area in Yellowstone National Park. Liu et al. (2008) found that combining DTs with multi-temporal Thematic Mapper images and GIS data was effective in determining spatial distribution and temporal changes of mangrove forests. The main issue with DTs is overfitting of the data, causing misclassification of pixels. To combat overfitting, pruning and boosting methods seek to improve the accuracy and misclassification (Pal and Mather, 2003).

1.2.4 Training Library

A spectral library is another approach to classification using a library of spectral response signatures of different land covers (Zomer et al., 2009). Spectral libraries are typically used in hyperspectral remote sensing, because the numerous bands allow for a more complete representation of the spectral reflectance signature of the target classes. Multispectral images have broad bands, which cover large parts of the electromagnetic spectrum, making spectroscopy difficult. Thus, they have historically not been used for spectral libraries. Not all spectral libraries use hyperspectral imagery. One study used multiple images to construct a library of values to classify images (Laborte et al., 2010). Laborte et al. used images from different time periods to improve the classification accuracy of a wetland in northern Laos. The authors found that classification accuracy strongly depended on the year and images used in the study. The major finding was that combining two images acquired from different seasons in the same year improved classification accuracy (Laborte et al., 2010).

1.3 Remote Sensing Imagery

Past work that utilizes remotely sensed datasets include the usage of satellites such as synthetic aperture radar (SAR), Landsat, and MODIS (Long et al., 2014; Rokni et al., 2014; Burke et al., 2016). MODIS data in particular has several different products that provide different functions. For example, MOD11A2 data provides composite eight-day thermal imagery while the MOD09A1 data provides an eight-day composite of daily optical reflectance data to produce one image with the best pixels. Several studies in the past have used MODIS as a way to quantify the extent of flooding and look at time-series of lake-surface areas (Amarnath and Rajah, 2015; Wang et al., 2014). MODIS data are not without limitations, with the main one being the coarse spatial resolution at 500m spatial resolution. This coarse resolution makes the

analysis of smaller study areas difficult but the problem is moot if the study area is large. Another issue is that with any optical remote sensing platforms, obtaining cloud free imagery can sometimes be a challenge. The only solution is to eliminate cloudy images from the analysis or use other bands, such as the thermal bands.

1.4 General Relevance of Research

The goal of this study is to develop a framework that better classifies flooding extent. To test this framework, I map the extent of flooding in the Chobe River Basin (CRB) during the years of 2014, 2015, and 2016. While there have been recent studies focusing on mapping flooding extent in the CRB, this study seeks to improve on those classification methods to better understand the flooding dynamic in this region. These floods have an impact on both the people and the vegetation in the region. The CRB has a vegetation composition of shrubs and grasses. Water availability affects rooting depths of these vegetation types. Decreasing water availability in El Nino years will cause the surface layer to become drier, which will increase the density of shrubs and other species that are able to access water in a deeper layer of the soil horizon (Pricope et al., 2015). In other words, perennial grasses and shrubs are influenced by flooding events, which translates to water availability as a substantial mechanism for grassland-to-shrub transitions. Additionally, if there is a decrease or increase in water availability, any type of human activity could be in jeopardy as well. Simply, if there is less water available due to El Nino and human alteration of land uses, then a change in vegetation composition is a possibility and one that needs investigation in terms of flood extent mapping.

The effects of El Nino and human alterations of land use on flooding extent are not well understood in the CRB. More information needs to be gathered to determine if flooding extent is decreasing or increasing, thereby changing vegetation dynamics and human population centers in the region. The results found here will also aid policy makers in making decisions regarding proper water resources and wildlife management, as well as how to manage human population centers. Finally, this research contributes to the advancement of knowledge of land cover classification using some of the newer techniques in remote sensing. Namely, support vector machines. This study also presents a new technique for classifying remote sensing imagery based on using a training library and splitting the study period into 3 time-windows. These

techniques seek to save the researcher time and effort, as well as attain a higher classification accuracy.

1.5 Research Objectives and Questions

The objectives of this study are three-fold. The first is to determine which classification technique is best at mapping the Chobe River basin using the training library. The second is to map inundation patterns in the Chobe River Basin using a training library and the time-window framework. The third is to quantify the extent of flooding in the CRB using the training library, with three classification algorithms; maximum likelihood, support vector machines, and decision trees. Additionally, my research will address the following questions:

1) Which classification technique produces the best classification results?

2) Can the use of a training library accurately classify wetland and open water pixels using separate time-windows for each image classification?

3) Has the magnitude of flooding extent in the CRB lessened or increased comparatively to previous years?

Chapter 2 - Study Area

2.1 Physical Landscape

The Chobe River Basin (CRB) is a seasonal marshland located on the border between Namibia and Botswana (Figure 2.1). The CRB is a sub-basin of the larger Upper Zambezi Basin with an area of roughly 4000 km^2 and a flat terrain with relief varying between 830 and 1050 m above mean sea level (Burke et al., 2016). The Zambezi River is the major river in the area, which originates in northwest Zambia, and collects rainfall from several countries, such as Zambia and Angola, before flowing through the study area and the CRB. Along with the Okavango Delta, the Chobe River is at the center of the Kavango-Zambezi Transfrontier Conservation Area (KAZA), which is the largest conservation area in the world, encompassing more than 170,000 km^2 and containing the world's largest elephant population (Burke et al., 2016). The KAZA provides a wildlife corridor, which contains water sources for agriculture and grazing. The Zambezi Wetlands and the Chobe River, which forms the southern border of the wetlands, are the main objects for classification.

The CRB has a heterogeneous mixed woody-herbaceous savanna ecosystem, with scrub woodlands in higher elevation zones and bushland, shrub, and thicket mosaics in lower elevations (Pricope et al., 2015). According to the world wildlife fund, the CRB is situated in three ecoregions; the Zambezian flooded grasslands, Zambezian and Mopane woodlands, and the Zambezian Balikiaea woodlands. The Zambezian flooded grasslands ecoregion covers a large part of the CRB, especially the Zambezi River part, and is a Flooded Grasslands and Savanna biome. This type of biome means that the soils are nutrient poor but the ecoregion provides habitats to a large number of animals because of the availability of food and water throughout most of the year ([March 1, 2017], World Wildlife Fund (panda.org)).

Soils in the region are thin alluvial and volcanic soils that have secondary grasslands situated on them for agricultural use. Specifically, the soils of the CRB are Entisols, which are soils of recent origin and are can be found in areas that have fluvial dynamics (USDA). Because the CRB is a seasonal wetland, this makes the Entisols here the subclass of fluvents. Fluvents are soils that have been deposited by the Zambezi to create a stratified layer of mineral and organic matter, although the organic matter decreases with depth (USDA).

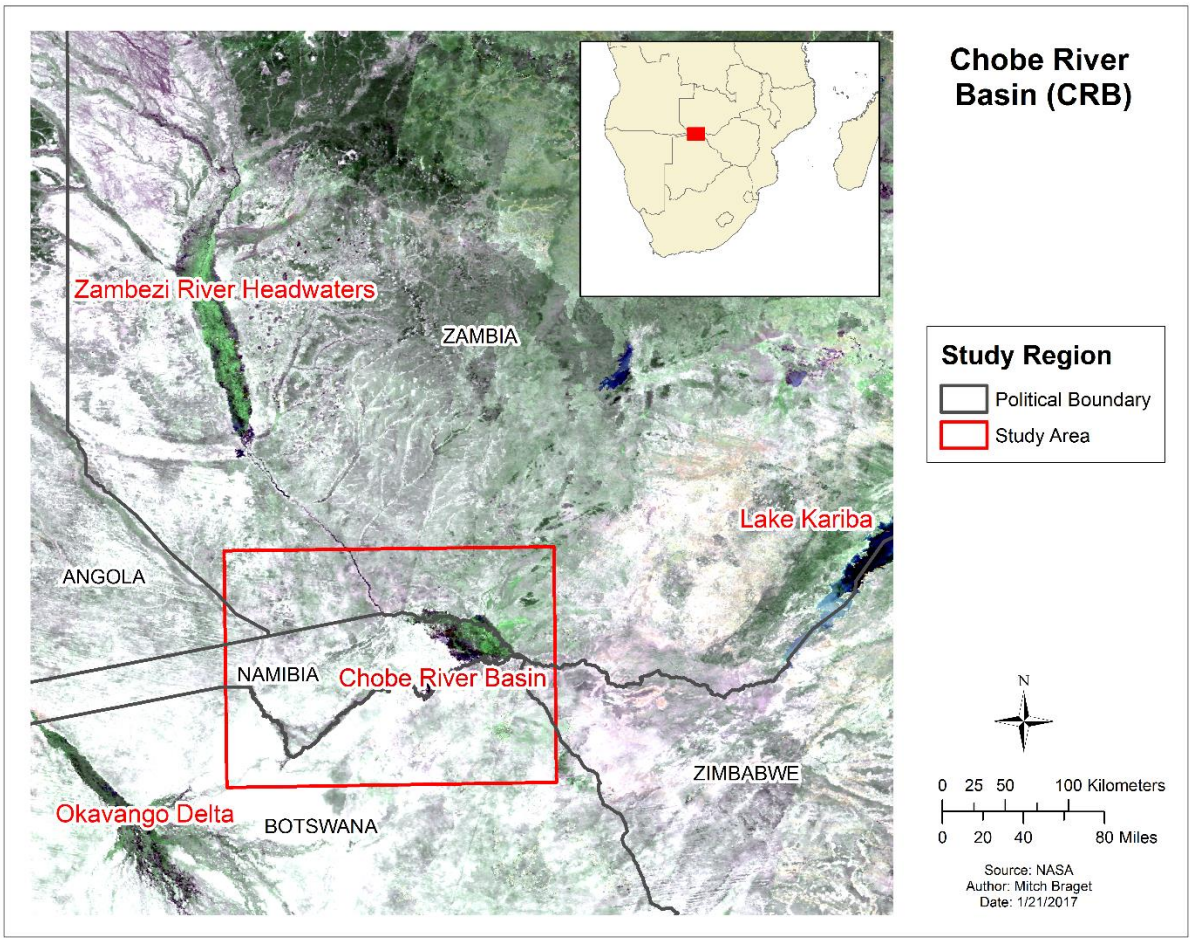


Figure 2.1 Study area (marked by the red rectangle) showing the CRB in Southern Africa, along with major topological features and Countries in the surrounding area.

Uplifting faults determine the shape of the CRB and the surrounding region (Figure 2.2). Faults are shown on the map with red lines. These faults influence the course of rivers and the blue lines influence the basins. Rivers flow into the basins and form floodplain wetlands, which gradually fill with sediment. The CRB exits over one uplifting fault on its eastern edge. This fault occurs in basaltic rock, which the Chobe and Zambezi rivers must erode down or else the water would backfill the CRB and create a lake.

Figure 2.3 displays the land cover in the CRB and comes from the Namibia Nature Foundation. Zambezi floodplain grassland dominates the area, which also happens to be the dominant ecoregion in the area. Along the Chobe River, the Chobe wetland land cover dominates the area. While not shown on this map, there are small communities scattered throughout the wetland, as well as the main city Kasane near the confluence of the Chobe and Zambezi rivers on the eastern part of the map (Burke et al., 2016).

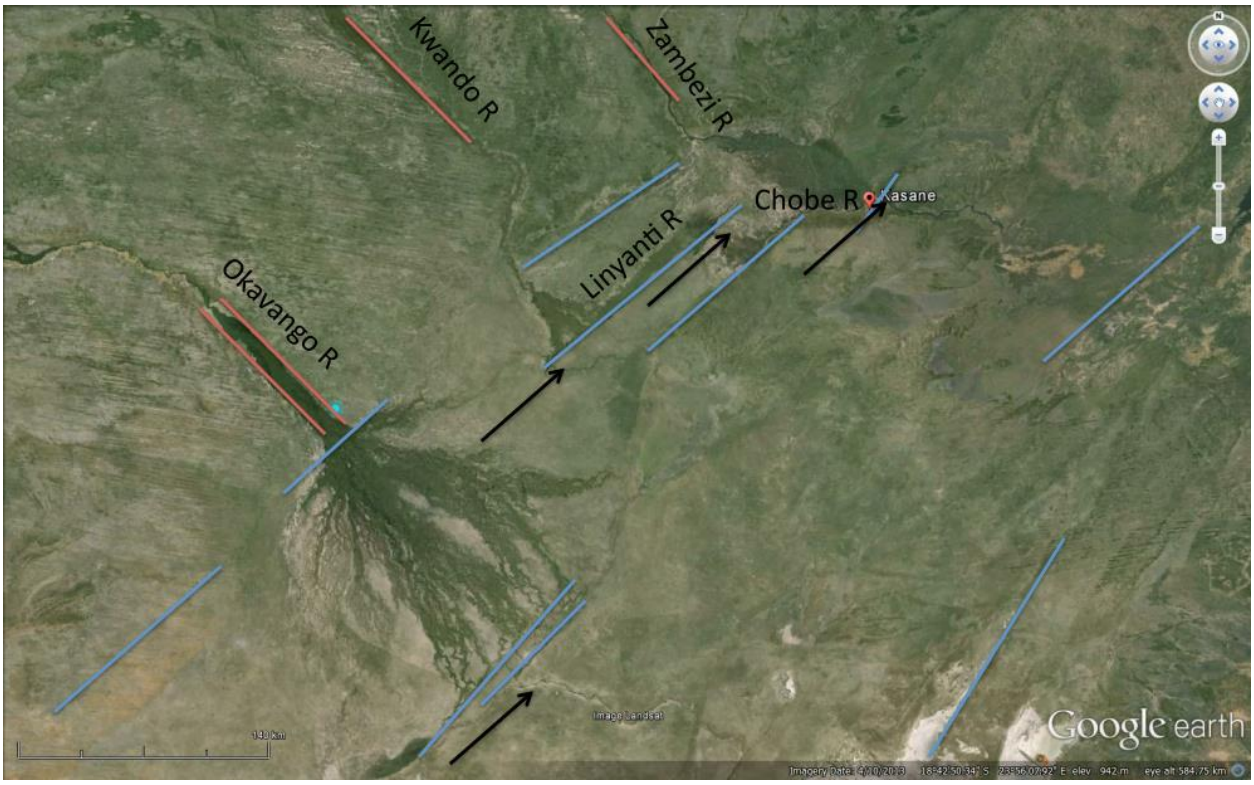


Figure 2.2 Blue lines show faults in a SW-NE direction and red lines show faults influencing river courses. (from Dr. Tooth powerpoint slide, 2016)

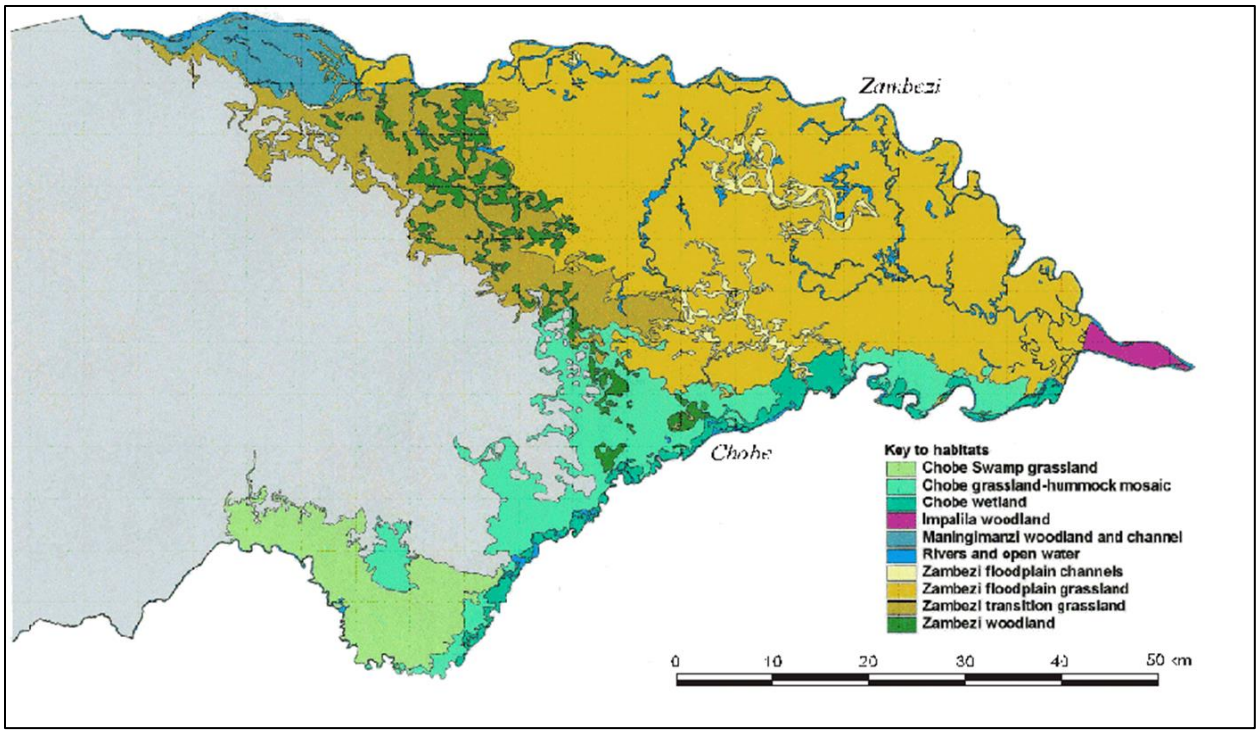


Figure 2.3 Land cover in the CRB according to the Namibia Nature Foundation

2.2 Population and Climate

The population in the CRB is approximately 90,000 as of 2011 (Burke et al., 2016). The settlements are communal conservancies that surround the Chobe River, with Kasane, located near the confluence of the Chobe and Zambesi rivers being one of the larger settlements in the area. These settlements are found near the floodplains because they use the water to implement flood recession agriculture or locally called Molepo farming. The climate is a steppe or semi-arid savanna (Koppen BSh) (Pricope et al., 2015). The entire Upper Zambezi catchment receives an average of 990mm of rainfall a year and the Kwando/Chobe sub-catchment, which provides water to the Chobe floodplain, receives around 800 mm/year of rainfall (Song et al., 2014). Most of the rainfall drops between November and April, which is the wet season for this area. The migration of the inter-tropical convergence zone (ITCZ) and the sea surface temperatures of the Indian and Atlantic oceans determine the regional precipitation (Burke et al., 2016).

2.3 Flood Pulses

The unusual property about the CRB is the flood pulses that occur from the Zambezi and the Linyanti rivers. A flood pulse occurs when there is a substantial amount of water going into a river system, water rises above the banks, and vegetation is inundated for periods of times ranging from a couple of weeks to months. Flood pulses bring in valuable nutrients and water that are vital to plant development and are a principal driving force for the productivity and types of biota in the river-floodplain region (Junk et al., 1989).

These pulse floods happen during the rainy season, which occurs between November and April. The floods consistently start at the end of February when the Zambezi flood pulse enters the Zambezi wetlands (Burke et al., 2016). Burke et al. also found that the best predictor of flooding extent in the CRB is the discharge of the Zambezi River 64 days prior to flooding. Thus, the peak of these floods occur at the end of the rainy season, which coincides with the end of April when the peak discharge of the Zambezi River reaches the CRB. When the Zambezi River overflows into the CRB, it reverses the flow direction of the Chobe River so that instead of flowing from west to east, it now flows from east to west (Pricope et al., 2015). The CRB also experiences a second flood pulse in the months of June and July because the Kwando River on the western side of the CRB has its peak discharge at that time of year, which is outside of the traditional rainy season. The water moves from the Kwando River, through the Linyanti

channel, then through Lake Liambezi, and into the CRB (Song et al., 2014). Therefore, the CRB has an unusual flood pulsing regime because it has two flood pulses during one year.

The flood waters often inundate large areas of the CRB but they do not persist year-round. As the flood waters recede towards the end of the rainy season, the Zambezi wetlands dry up, with small pockets of wetlands staying year round (Pricope 2012). The lower floodplain of the Chobe-Linyanti contains standing water until the middle of the dry season. The remaining standing water that persists year-round include Lake Liambezi, the Chobe and Zambezi channels, portions of the northeastern Zambezi wetlands, and the area of the Zambezi wetlands near Kasane (Burke et al., 2016).

Chapter 3 - Comparison of Classification Algorithms

To address the first objective of which decision rule is best for classifying land cover in the Chobe River basin, I evaluated the decision rules of Maximum Likelihood classifier (MLC), Decision Trees (DTs), and Support Vector Machines (SVMs). To test the various rules under a variety of conditions, imagery was selected from the start, middle, and end of the flooding season. This chapter first highlights the image acquisition and preprocessing stage. The Kauth-Thomas tasseled cap transformation as the feature selection and a short section on accuracy assessment follows. The chapter concludes with the results of the three decision rules and a discussion of which one produced the highest overall accuracy and kappa values. Reasons for each decision rule performing satisfactory or poorly are discussed.

3.1 Methods

3.1.1 Image Acquisition and Preprocessing

For this analysis, MODIS MOD09A1 imagery was acquired from the LP DAAC website (https://lpdaac.usgs.gov/). The spatial resolution of this dataset is 500m and consists of seven reflectance bands (table 3.1). Each composite consists of compiling daily images over eight days and pixels are selected based on quality, cloud, and viewing geometry, until the highest quality single value per pixel remains to use in the composite (Chen et al., 2013). Imagery was collected from early February to early July for the years 2014, 2015, and 2016. Table 3.1 breaks down the number of images for each year. A total of 59 composites were collected by the MODIS sensor during the analysis time period but only 49 were used in the analysis due to clouds. A total of ten images were excluded from the classification due to excessive cloudy conditions, with four in 2016, three in 2015, and three in 2014. The period of data collection was chosen because the rainy season coincides with this time period and it is an objective to quantify the flooding extent in the CRB. The 8-day composite data were re-projected from a sinusoidal projection into UTM 34S using the MRTWEB tool (USGS). Each band in this MODIS dataset consists of 16-bit signed integer reflectance values that range from -100 to 16000. All bands in each composite were multiplied by 0.0001 to rescale to true reflectance. Each of the seven bands for a given date were stacked into one image. The study area was subset to the CRB based on studies by Pricope (2012).

Table 3.1 Breakdown of MODIS imagery used per year for classification

Year	Images Used	Description	Projection
2014	16	500 m spatial resolution, bands 1-7	UTM 34S
2015	17	500 m spatial resolution, bands 1-7	UTM 34S
2016	16	500 m spatial resolution, bands 1-7	UTM 34S

3.1.2 Calculation of Kauth-Thomas Components

Following preprocessing, each image was converted into the Kauth-Thomas tasseled cap transformation using a Python script in ArcGIS (Environmental Systems Research Institute, Redlands, California). The tasseled cap transformation rotates the data in spectral feature space using a linear combination of inputs from image spectral bands. It is primarily used to enhance biophysical properties in the imagery by creating, synthetic components corresponding to the Brightness, Greenness, and Wetness properties of the data (Lobser and Cohen, 2007). It thus converts raw reflectance values into a few, more easily interpretable components, which can be used as input into classification algorithms (Kauth and Thomas, 1976). It was originally developed to track the development of seasonal agricultural crops, and has subsequently been used to temporal analysis of surface biophysical characteristics. The Kauth-Thomas tasseled cap transformation was used because it transforms data into component bands that represent biophysical properties and decreases dimensionality of the data.

The K-T components for each image were calculated by multiplying each band of the MOD09A1 composite was multiplied by coefficients derived for the MODIS instrument by Lobser and Cohen (2007). These coefficients are shown in table 3.2 and equations [3.1], [3.2], and [3.3]. Tasseled cap coefficients are instrument specific and time sensitive, as each Landsat sensor has their own set of coefficients.

Table 3.2 MODIS tasseled cap coefficients

	Red (B1)	NIR1 (B2)	Blue (B3)	Green (B4)	NIR2 (B5)	SWIR1 (B6)	SWIR2 (B7)
Brightness	0.4395	0.5945	0.246	0.3918	0.3506	0.2136	0.2678
Greenness	-0.4064	0.5129	-0.2744	-0.2893	0.4882	-0.0036	-0.4169
Wetness	0.1147	0.2489	0.2408	0.3132	-0.3122	-0.6416	-0.5087

NIR, near infrared and SWIR, short-wavelength infrared

$$Brightness = (Kb * B1) + (Kb * B2) + (Kb * B3) + (Kb * B4) + (Kb * B5) + (Kb * B6) + (Kb * B7) \quad (\mathbf{3.1})$$

$$Greenness = (Kg * B1) + (Kg * B2) + (Kg * B3) + (Kg * B4) + (Kg * B5) + (Kg * B6) + (Kg * B7) \quad (\mathbf{3.2})$$

$$Wetness = (Kw * B1) + (Kw * B2) + (Kw * B3) + (Kw * B4) + (Kw * B5) + (Kw * B6) + (Kw * B7) \quad (\mathbf{3.3})$$

Where: Bx = reflectance in MODIS band x

Kx: = Kauth-Thomas coefficients for brightness, greenness, or wetness

3.1.3 Land Cover Classes

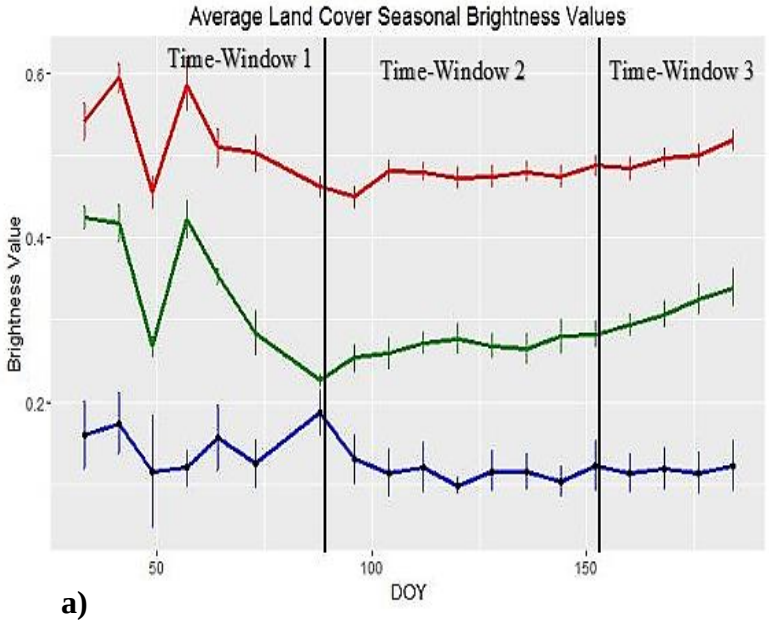
As noted in the methodology background section, three land cover classes; open water, emergent vegetation, and non-water, were used in this classification. Defining the land covers required some knowledge from local sources. Photographs taken in the study area were used to define the characteristics of the three land cover categories to identify areas representative of each type. Open water (Figure 3.1) contains visible water surface with little to no vegetation coverage. Emergent vegetation is defined here as a mixture of water and aquatic vegetation (Figure 3.2). Non-water is composed of the remaining land covers, such as shrubland, rangeland, etc. Although other land cover classes could be defined for the study area, adding more classes to the analysis would deemphasize the focus of mapping the flooding in the region.

The open water and emergent vegetation land covers were kept separate because the emergent vegetation class contains a mixture of both open water and vegetation overlying water or saturated soil. In effect, it resembles marshland (Mitsch and Gosselin, 1993). Open water (blue) and emergent vegetation (green) also exhibit different tasseled cap component values in imagery (Figure 3.4), which if combined together could result in misclassification. Non-water (red) is also spectrally different from emergent vegetation and requires its own class. It is important to note that the classes are transiently defined based on current conditions. At certain times of the year, a given area is emergent vegetation and as the flooding occurs, it transits into an open water class. Some areas are persistently one land cover, where an area consistently represents the open water class, such as lakes and river channels. Basically, the emergent vegetation and open water classes define the conditions for an area that is flooded.

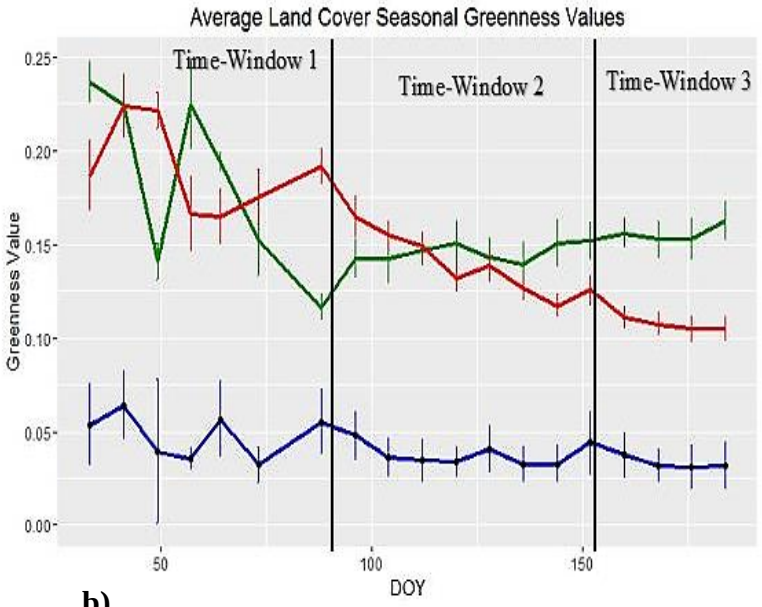
Figure 3.1 Typical landscape in the Chobe River basin during a flooding event. This photograph is representative of the open water class. (Courtesy of Dr. Stephen Tooth, Aberystwyth University)

Figure 3.2 The emergent vegetation land cover in the Chobe River basin. This photograph is representative of the emergent vegetation class. (Courtesy of Dr. Stephen Tooth, Aberystwyth University)

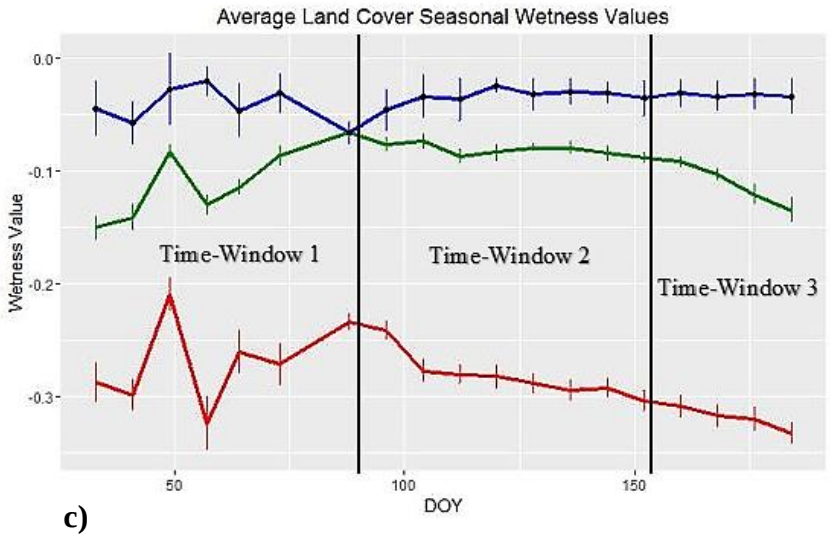
Figure 3.3 The non-water land cover in the Chobe River basin. This photo is representative of the non-water class. (Courtesy of Dr. Stephen Tooth, Aberystwyth University)



a)



b)



c)

Figure 3.4a-c Seasonal tasseled cap component values for open water (blue), emergent vegetation (green), and non-water (red), from 2014 to 2016. Figures 3.4a, b, and c display the brightness greenness, and wetness component, respectively, land cover values through the study period.

3.1.4 Classification Procedures

I chose MLC because it has been widely used in the literature as a base classification algorithm to compare newer techniques to and serves the same purpose here. While MLC consistently provides lower accuracy results compared to other techniques, it offers an initial classification result to compare against. I chose DTs and SVMs because they are more recent techniques, are non-parametric classifiers, and fundamentally use the feature space differently than MLC. DTs and SVMs also do not get stuck in local minima such as Neural Networks. DTs are simple to understand and interpret because the trees can be visualized. DTs were also picked because the arrangement of training samples in feature space suggested that DTs could easily partition the data and result in high accuracy values. SVMs work well with small amounts of training samples and are easily understood compared to Neural Networks. SVMs can also use kernel functions to increase accuracy results. Overall, the three techniques described here were the most convenient for classifying imagery. In the following subsections, a brief explanation of the three classification algorithms is provided.

Maximum Likelihood Classifier

MLC is a parametric supervised classification algorithm that uses the means and variances of each training class to compute probability distribution functions and assign a pixel to a class (Srivastava et al., 2012). The MLC algorithm uses the Bayes' theorem for its decision making, in which pixel values of each class in the multidimensional feature space are normally disturbed. The assumption of normally distributed data means the mean vector and the covariance matrix characterize a class. Using the vector and covariance of each pixel, the probability is calculated for each class to determine the membership of each pixel to the class (Pal and Mather, 2003). The pixel is classified to a class to which it has the highest probability of being associated with that class. The Bayes' theorem in which the weighted distance or likelihood D of unknown measurement vector X belonging to a known class, M_c, is based on the Bayesian equation [3.4] (Otukei and Blaschke, 2010),

$$D = ln(Ac) - [0.5ln(|covc|)] - [0.5(X - Mc)T(covc - 1)(X - Mc)] \quad \textbf{(3.4)}$$

Where:

D = weighted distance (likelihood),

c = a particular class,

X = the measurement vector of the candidate pixel,

M_c = the mean vector of the sample of class c,

Ac = percent probability that the current pixel is a member of class c,

Cov_c = the covariance matrix of the pixels in a sample class c,

$|Cov_c|$ = determinant of Covc,

Cov_c^{-1} = inverse of Covc,

ln = natural logarithm function,

T = transposition function.

A decision surface using MLC is determined by shapes of the surfaces which separate one class from another in feature space. Each class is defined by regions in multispectral space where their discriminant functions [4] are the largest. Figure 3.5 demonstrates two probability density functions overlapping. So for example, unknown pixel-vector X would be assigned the class whose probability density is greater than that of any other class. Using the Kauth-Thomas components as the input for classification, MLC was run using ENVI (Exelis Visual Information Solutions, Boulder, Colorado).

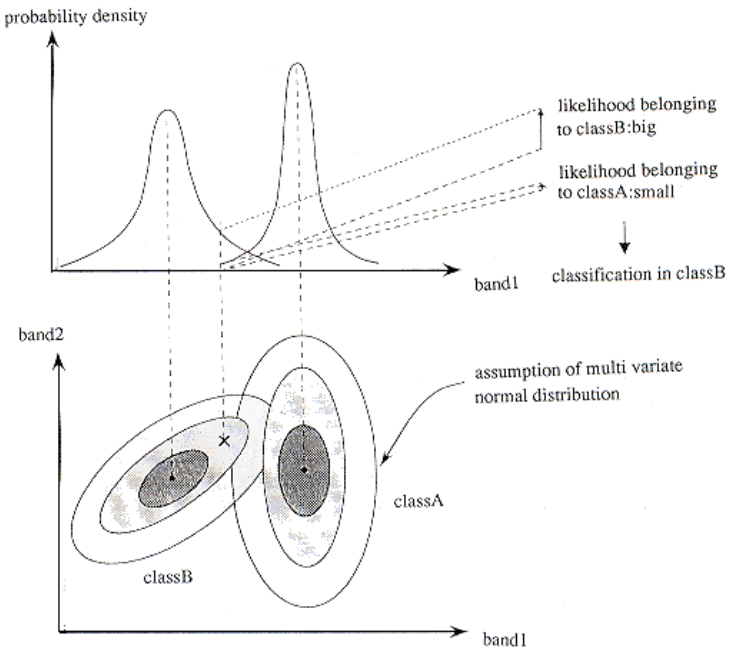


Figure 3.5 Each contour represent equal probability membership of some class

Decision Tree

Decision trees (DTs) are a non-parametric supervised classification technique, which makes no assumptions about the distribution of training data. DTs are built by having one root

node, several internal nodes, and finally the terminal nodes, which are the final land covers (or other class property) (Outkei and Blaschke, 2010). Training data moves down the DT based on a user-defined classification framework until a terminal node is reached, classifying the pixel into a class. This framework is considered a chain of simple decisions rather than a single, complex decision other supervised classification techniques use (e.g. Support Vector Machines). A DT partitions data down a tree until a pixel reaches a terminal node or class (Figure 3.6).

Decision trees are derived by successive splits of data based on a series of binary decision properties. Each split is based on a threshold, expressed in one dimension (Figure 3.6). The threshold values are set to maximize the amount of information that's contained in each new split. A number of methods to split nodes can be used including the Gini Index, information gain, entropy, and standard error. The Gini index measures the change of a randomly chosen element from the training data would be mislabeled if it was randomly labeled according to the distribution of labels (Waske and Van der Linden, 2008). The information gain ratio determines every split that it can make and picks the one that gains the most new information (Fawagreh et al., 2014). Entropy is a way of measuring how much internal information there is in each split and quantifies the amount of information gain to determine the split that adds the most information (REF). In this application, the standard error was used for splitting each node.

During the construction of the trees, a problem that can occur is overfitting of the data, where the decision rule attempts to incorporate one instance of data that does not occur near a cluster of its own class. The overfitting results in misclassification of instances due to the classifier incorporating more data of another class in its attempt to collect one instance. Pruning is used to eliminate the effects of overfitting by removing section of the tree that have little contribution to classifying instances (Fawagreh et al., 2014).

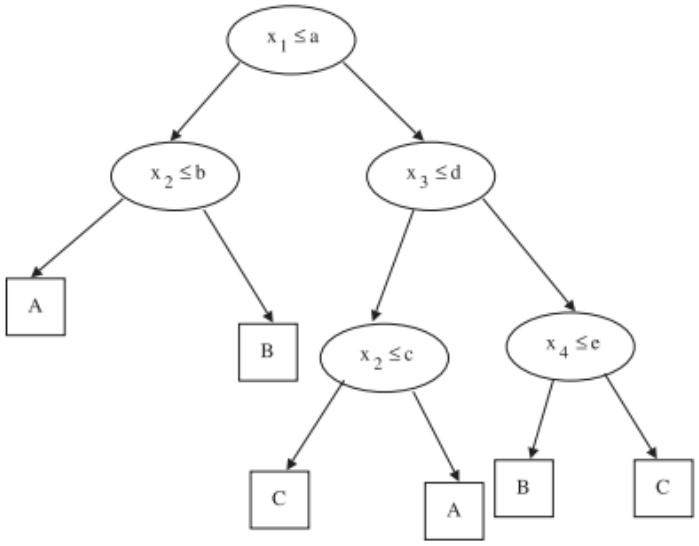


Figure 3.6 Example of a decision tree, where X_i are feature values; a, b, c, d, and e are the threshold and A, B, and C are class labels (source: Pal and Mather, 2003)

DT's partition the data in feature space (Figure 3.7). A univariate DT (Figure 3.7) is where decision boundaries at each node in a tree are defined by the result of the test applied to a single feature and is evaluated at each internal node (Pal and Mather, 2003). A multivariate DT uses a linear combination of features when it splits the data into nodes (Pal and Mather, 2003). Simply, multivariate splits incorporate two variables at each node to determine the best split, whereas univariate only uses one variable to split at the node. A univariate approach was used with the three Kauth-Thomas transformation bands of brightness, greenness, and wetness as the decision rules. Values of these bands determine how data is partitioned. A univariate DT is used instead of a multivariate DT because only one band is used at each node to partition the data.

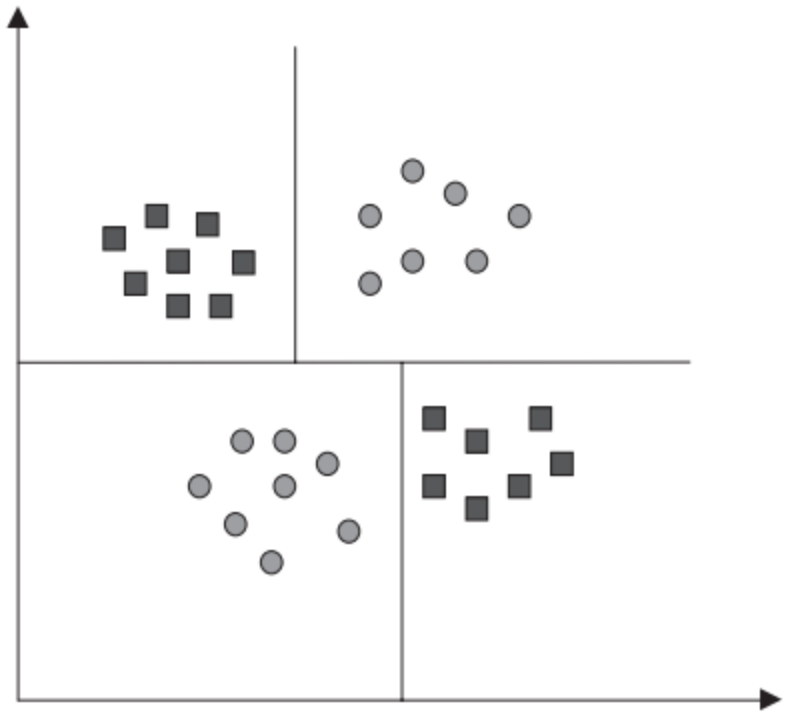


Figure 3.7 Axis-parallel decision boundaries of a univariate decision tree (source: Pal and Mather, 2003)

The DTs were made in the R statistical computing software using the mvpart package, with the training library as the input. A DT was made for each time-window for each year of the study, with a total of 9 DTs. Figure 3.8 is an example of a DT created for time-window 2, 2014. Pruning was run on each DT but no significant change in partitioning values was noted. Each tree was constructed and run in ENVI to create the final classified imagery.

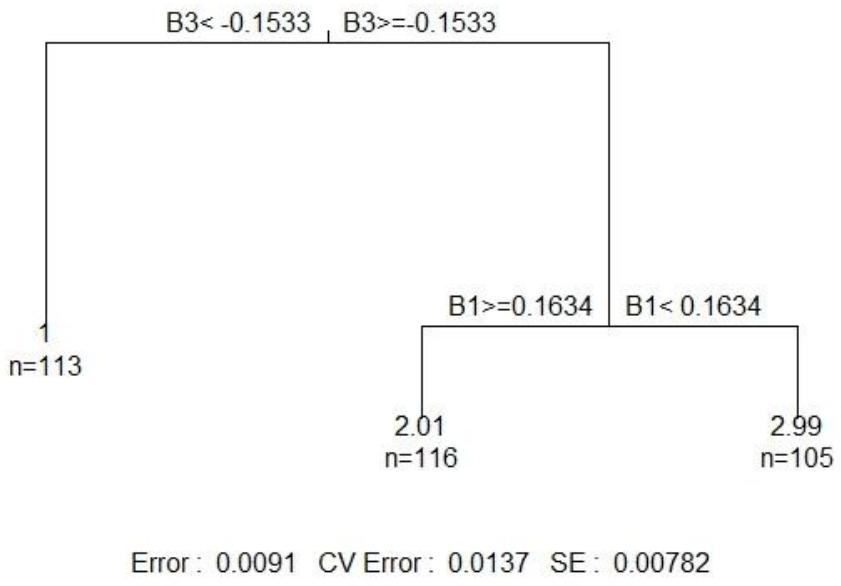


Figure 3.8 Example DT in this analysis for Time-Window 2, 2014 created in R statistical programming language

Support Vector Machines

Support Vector Machines (SVMs) are a non-parametric, supervised statistical machine learning technique that makes no assumptions about the data distribution. SVMs determine an optimal hyperplane in a set of data by using support vectors to maximize the margin or distance between the support vectors and the hyperplane (Waske and Van der Linden, 2008). The optimal hyperplane is determined through an optimization problem using Lagrange multipliers and quadratic programming (Srivastava et al., 2012). Many hyperplanes may exist in feature space, but the optimal hyperplane minimizes classification errors and the plane represents the two classes (Figure 3.9). There are two supporting hyperplanes on which the support vectors are located to help maximize the distance from the data points of each class (Figure 3.6) (Adam et al., 2014). The support vectors represent the training samples toughest to classify and help to identify the maximum margin for the hyperplane and classification of the data (Mountrakis et al., 2011). Hyperplanes may not be able to separate data due to points with large errors. A soft-margin hyperplane, introduced by Cortes and Vapnik (1995), incorporates slack variables (ξ), which represent the distance of incorrectly classified data from the hyperplane (Kavzoglu and Colkesen, 2009). The slack variables (Figure 3.10) incorporate misclassified data points in

feature space by introducing the penalty parameter. The penalty parameter C, and to some degree the slack variables, controls the shape of the SVM by allowing for some misclassification of the data (Waske and Van der Linden, 2008). Data with high penalties can lead to overfitting and lower accuracy results.

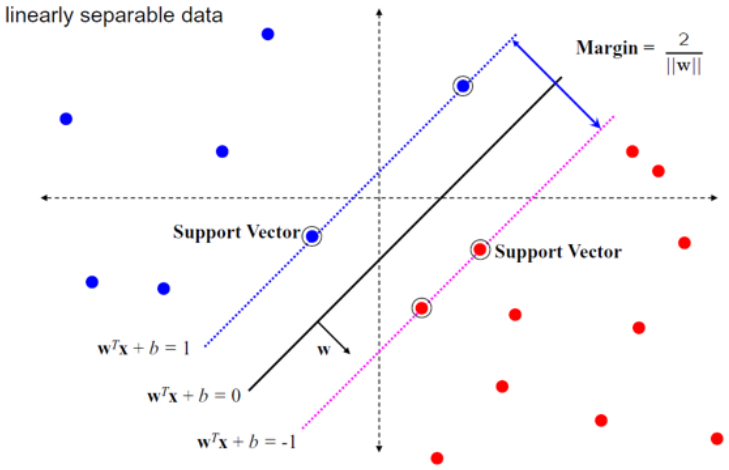


Figure 3.9 Support vectors (circled) define the margin of the hyperplane (source: Dr. Shawn Hutchinson, Geocomputation)

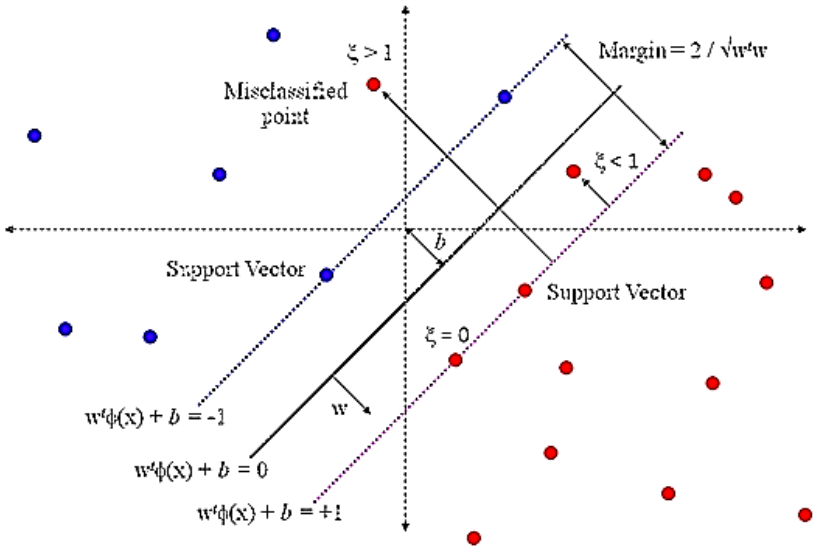


Figure 3.10 Nonlinear data sets and solution by introducing slack variables (source: Dr. Shawn Hutchinson, Geocomputation**)**

Not all data is linearly separable and SVMs can be optimized to search for a non-linear hyperplane using a kernel function. The kernel function maps the data into a higher-dimensional feature space and decreases the computational demand (Kavzoglu and Colkesen, 2009). The

kernel function reworks the optimization of the hyperplane with the nonlinear mapping function Φ and finds the optimal hyperplane to separate the data. The optimization of the hyperplane to use Φ involves Legrange multipliers to equation [3.5], to produce the final result in equation [3.6], where α_i is the Legrange multipliers, x_i is a given training case with a class membership of (y_i)* α_i, and K(x, x_i) is the kernel function (Kavzoglu and Colkesen, 2009). Also known as the kernel trick, equation 6 does not require knowing Φ to map data in the newly transformed feature space (Figure 3.11) (Waske and Van der Linden, 2008).

$$\left(\Phi(x_i)\Phi(x_j\right) = k\left(x_i, x_j\right) \quad \textbf{(3.5)}$$

$$f(x) = \sum_{i=1}^{L} \left(a_i\, y_i \mathrm{k}(x_i, x_j + b\right) \quad \textbf{(3.6)}$$

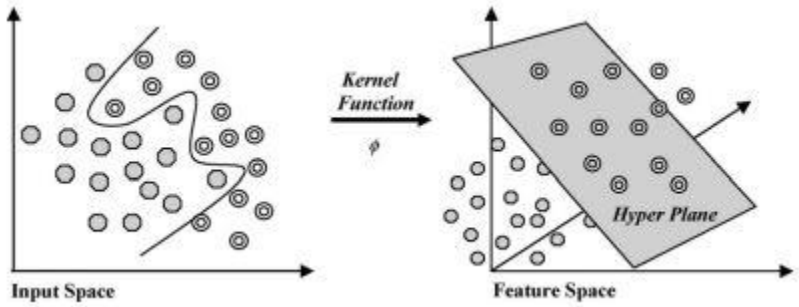


Figure 3.11 Mapping data into a high-dimensional feature space with a Kernel Function (source: Kavzoglu and Colkesen, 2009)

Common kernel functions used in SVMs include the linear, sigmoid, polynomial, and the Gaussian radial basis function. The polynomial and radial basis function kernels (RBF) are commonly used in remotely sensed data (Adam et al., 2014; Otukei and Blaschke, 2010; Srivastava et al., 2012). Equations [3.7] and [3.8] show the polynomial and RBF kernels, respectively, where α denotes the order of the polynomial for equation [3.7], and the γ and ||x|| in equation [3.8] are the gamma or kernel width for the RBF and the norm of x. The RBF kernel was used because a number of studies have found it to produce high accuracies compared to the polynomial kernel (Kavzoglu and Colkesen, 2009; Adam et al., 2014; Mountrakis et al., 2011).

$$k\left(x_i, x_j\right) = \left(x_i * x_j + 1\right)^a \quad \textbf{(3.7)}$$

$$k\left(x_i, x_j\right) = exp\left[-\gamma \left|\left|x_i - x_j\right|\right|^2\right] \quad \textbf{(3.8)}$$

The RBF kernel requires two parameters, C and gamma (γ). As mentioned previously, the C parameter is the penalty parameter that assigns a penalty to misclassified data points in feature space and sets the importance of maximizing the margin between the hyperplane and the support vectors while minimizing the amount of slack variable values (Ben-Hur and Weston, 2010). The gamma parameter (γ) determine the kernel width or margin definition of the hyperplane (Kavzoglu and Colkeson, 2009). Varying the two parameters has an effect on the shape the kernel takes in feature space. A decrease in γ will decrease the curvature of the decision boundary (Figure 3.12), while an increase in C forces the curve to accommodate for the larger penalty (Ben-Hur and Weston, 2010). Increasing just γ parameters results in overfitting of the data around the white areas (Figure 3.13).

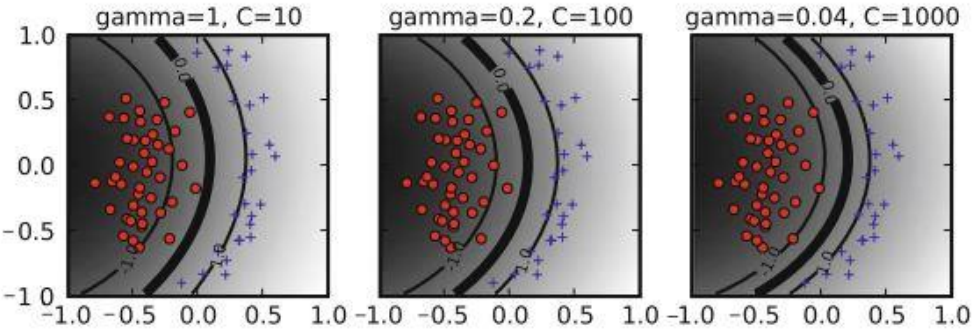


Figure 3.12 Variation of the parameters gamma (γ) and C (source: Ben-Hur and Weston, 2010)

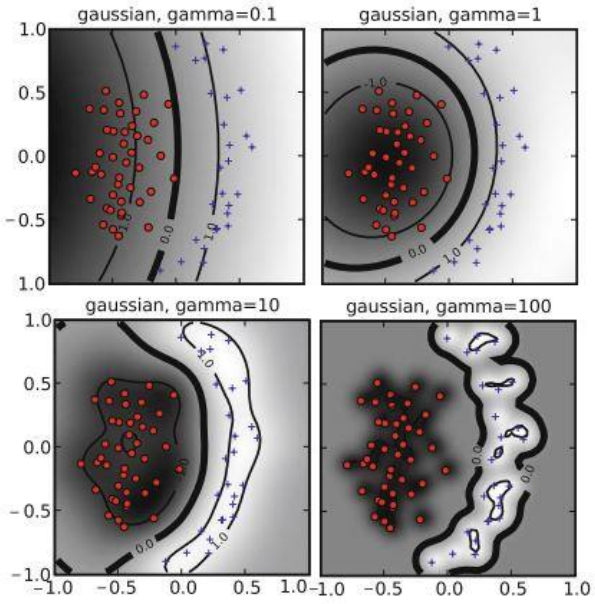


Figure 3.13 The effects of varying the gamma (γ) parameter (source: Ben-Hur and Weston, 2010)

The LIBSVM library was used to select the optimal parameters for the RBF kernel. LIBSVM uses a V-fold cross validation grid search on the training data to optimize the C and γ parameters (Chih-Chung and Lin, 2011). The V-fold cross-validation divides the training data into v subsets of equal size and then one subset is tested using a classifier training on the remaining v-1 subsets, allowing the whole training set to be predicted (Ben-Hur and Weston, 2010). The goal is to find the pair of parameters with the highest cross validation accuracy. Figure 3.14 demonstrates a grid search with the x-axis as the gamma value and the y-axis the C parameter. The values to the right show the accuracy values associated with the combination of different parameters. To implement a multiclass-based SVM model, the one-against-one procedure is used (Adam et al., 2014). Support vector machine classification was implemented using an IDL script and the ENVI software.

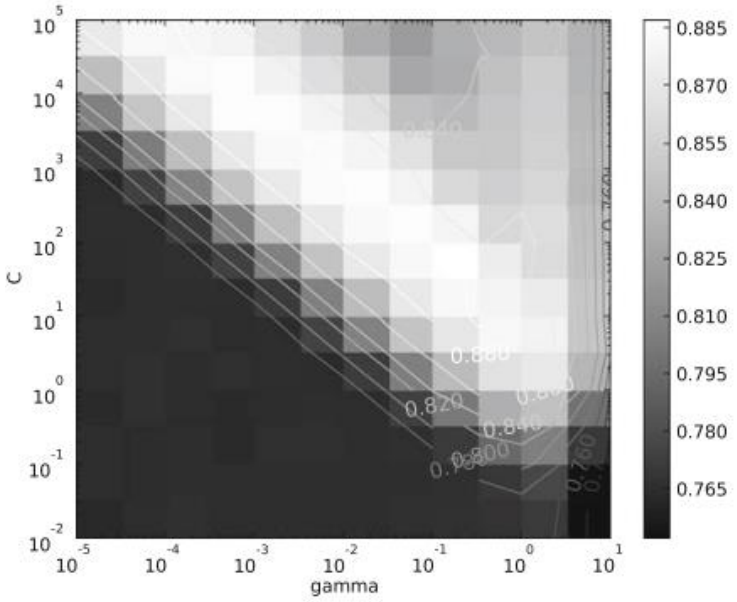


Figure 3.14 Grid search of parameters C and gamma (γ) (source: Ben-Hur and Weston, 2010)

3.1.5 Accuracy Assessment

Landsat 8 OLI 30 m spatial resolution true-color imagery closest to the classification date was used to evaluate the thematic maps developed from the classification algorithm (table 3.3). Research has shown that using reference data that is of higher resolution than the map classification is an acceptable accuracy assessment procedure (Olofsson et al., 2014). An error matrix was constructed to assess the overall accuracy, user's and producer's accuracies, and the kappa statistic. All three land covers were visually interpreted in the Landsat 8 OLI imagery. A

stratified disproportionate random sample of 30 to 50 validation points was generated for each of the three land covers for all 27 reference images in the validation process. The sample size for the accuracy assessment was determined from past classification accuracy research articles (Olofsson et al., 2014). Non-water contained the least amount of validation points because misclassification of non-water values was rare and more points inflated the accuracy of the classification. A total of 27 error matrices were created. Each classification technique consisted of three matrices per year, 1 for each time-window, for a total of 9 over the study period, with three techniques in total. The kappa statistic measures the difference between the observed agreement between reference data and the classifier used to complete the classification versus the possibility of agreement between reference data and a random classifier (Adam et al., 2012). Overall accuracy represent the probability that a randomly selected point is classified correctly on the map. Producer's (errors of omission) and user's (errors of commission) accuracy represent the probability of a reference pixel being classified correctly and the probability that a pixel classified on the map representing that category on the ground, respectively (Olofsson et al., 2014).

Table 3.3 Nine Landsat 8 OLI reference images

Date	Path	Row	Description	Projection
17-March-2014	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
20-May-2014	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
21-June-2014	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
20-March-2015	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
23-May-2015	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
24-June-2015	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
19-Feburary-2016	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
25-May-2016	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S
26-June-2016	174	72	30 m spatial resolution, bands 2 - 7	UTM 34S

3.2 Results of Classification Techniques

The results of running the three decision rules are presented here and help to address the first question of the study, which is to determine the best decision rule to map the flooding extent in the CRB. In terms of OA and kappa over the entire study period, from 2014 to 2016, and incorporating all three time-windows, SVMs and DTs provided virtually the same performance (table 3.4). SVMs provided the highest performance the years 2015 and 2016, while DTs had

the highest performance for 2014. MLC consistently had the lowest OA and kappa values across individual time-windows, single years, and over the entire study period. Confusion of the open water and emergent vegetation classes contributed the most to the lower accuracies produced by MLC. Both classes closely resemble one another in terms of tasseled cap component values. Clouds also caused overall classification confusion, especially during time-window 1 for the study period. Clouds were classified as part of the emergent vegetation class for all three techniques, requiring cloudy images to be eliminated from the analysis. Overall, there were slight differences in kappa and OA but SVMs and DTs were virtually the same when it came to mapping flooding extent in the CRB.

Table 3.4 Average OA and Kappa for the three decision rules for a given year and for the study period of 2014 to 2016

	2014	2015	2016	2014-2016
All Techniques OA	0.84	0.89	0.86	0.86
All Techniques Kappa	0.76	0.84	0.79	0.79
DT OA	0.86	0.90	0.86	0.87
DT Kappa	0.80	0.85	0.78	0.81
SVM OA	0.84	0.91	0.88	0.88
SVM Kappa	0.76	0.87	0.83	0.82
MLC OA	0.80	0.86	0.84	0.84
MLC Kappa	0.71	0.79	0.76	0.75

3.3 Discussion of Classification Techniques

Classification performance depends on a number of factors, not only the training data and classification algorithm. The classification scheme, pixel spatial resolution, study area characteristics, and quality of reference data contributes to the accuracy in any study (Shao and Lunetta, 2012). The accuracy assessment shows fairly high OA and kappa results (table 3.4) for DT and SVMs across both the time-windows and for entire years. The first time-window proved to be the most difficult to classify due to an abundance of cloud cover. This cloud cover resulted in land misclassified as emergent vegetation (Figure 3.15). Overall, the majority of the misclassification came from open water and emergent vegetation confusion. A reason for these misclassifications is due to the validation procedure. Because the validation imagery was 30m Landsat 8 data and the classification imagery was 500m MODIS data, the confusion of land

cover is due to multiple cover types existing in the same pixel, forcing the classification rule to decide which land cover to classify the pixel as. For example, a 500m pixel has a mixture of open water and emergent vegetation. The algorithm decides to classify the pixel as an emergent vegetation pixel, but during the validation process, the researcher sees open water as the dominant land cover in the area. Another reason for the mix-up between open water and emergent vegetation is due to the similar spectral signatures being relatively close to one another. The open water and emergent vegetation classes have similar values around March 29 and then start to disassociate afterwards (Figure 3.16). The mix-up between these two classes led to an overestimation of the open water class in some of the imagery that occurred in time-windows 1 and 2.

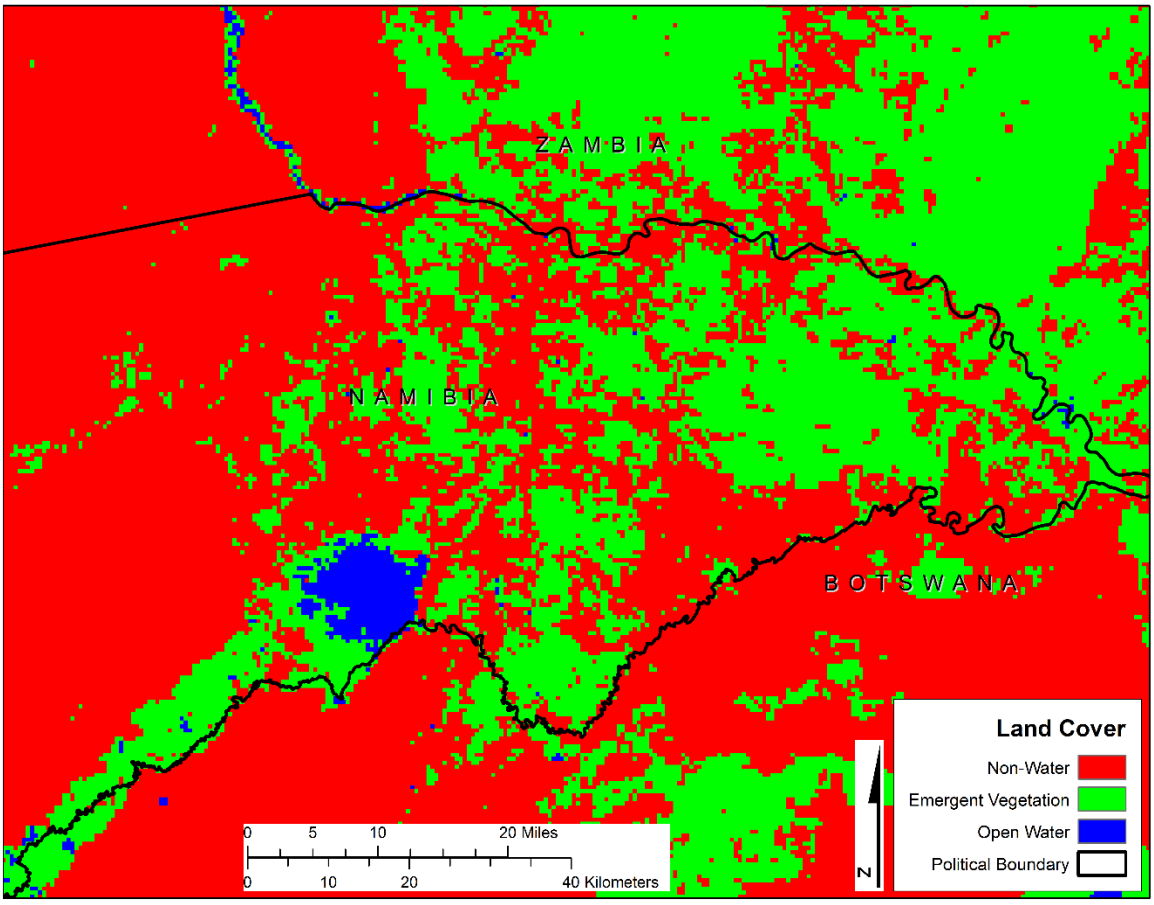


Figure 3.15 February 2, 2015 classified imagery using DT

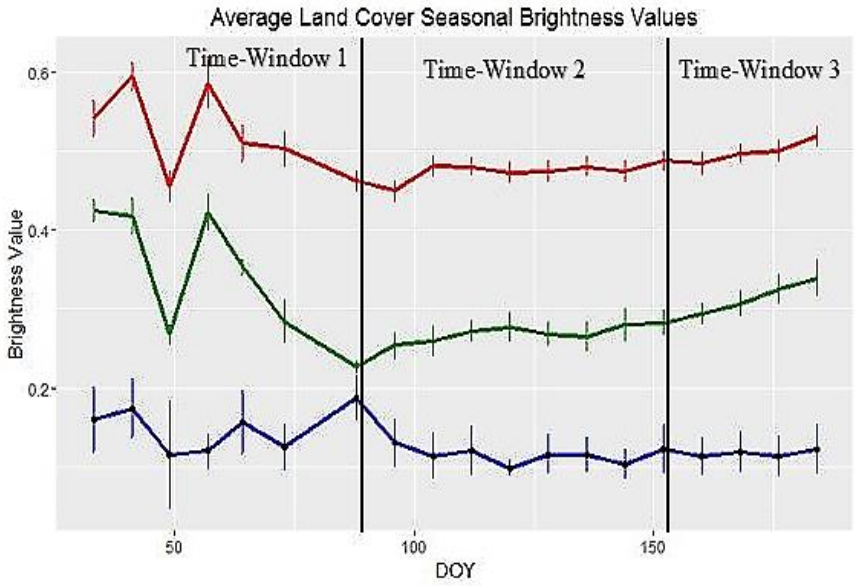


Figure 3.16 Open water (blue) and emergent vegetation (green) values from the brightness band

Not all of the misclassification between the two classes is explained by the tasseled cap component values of the emergent vegetation and open water land covers. Some of the misclassification is due to the researcher's sampling procedure. Namely, there was roughly an equal number of training pixels, 50, acquired for each class. The imagery itself did not contain equal areas of the three land covers throughout the year. For example, during the second time-window, there were more values of open water and emergent vegetation than land pixels. Therefore, providing more land pixel values for a classification technique would have reduced what Wright and Gallant (2007) called class imbalance. Class imbalance is when you over represent a class which does not have as many pixels in the imagery as other classes (Wright and Gallant, 2007). In the future, a more thorough sampling procedure would produce higher accuracy between the open water and emergent vegetation land covers but it would not be a substantial increase.

In terms of classification techniques, DTs and SVMs had comparable accuracy across each of the 3 time-windows. SVMs were best for the entire year for 2015 and 2016, while DTs were best in 2014, DTs had the higher overall accuracy and kappa values for time-windows 1 and 3and SVMs were best for time-window 2. Therefore, both of these classification algorithms produced high accuracies and it is difficult to say which one is best to use in future studies. From a practical standpoint, the setup and application of SVMs was difficult to implement, with

a roundabout method in ENVI being the only way to use the training library. SVMs also required two user-defined parameters and further increased the chances of classification error. Though, SVMs did produce higher accuracies across a given year. Meanwhile DTs were very simple to implement, and were able to produce higher accuracies in two of the three time-windows. These two time-windows were the more rapidly changing time-windows in terms of changing tasseled cap component values of the three land covers. SVMs were best for time-window two, which is the most stable of the three time-windows (Figure 3.2).

A possible reason the SVM classifier performed poorly for an individual time-window could be because SVMs are sensitive to noisy data and outliers (Mountrakis et al., 2011). Using kernel functions help to classify nonlinear data by plotting the data into another space. The C parameter can apply a penalty parameter for misclassified pixels, and further decrease the chances of misclassification. However, if there is a large number of outliers in the training data, then the penalty parameter can start to include pixels that already belong to their right land cover, further causing misclassification error (Ben-Hur and Weston, 2010). Another reason for SVMs producing lower accuracy values than DTs could be because of the two user-defined parameters. The C and γ derived from LIBSVM switches between scaled and unscaled data because visually the scaled data would produce the best results and other times the unscaled data would produce the best results (table 3.5). The grey parameters were the ones used in the SVM classification. The parameters lack consistency moving from time-window to time-window and could be decreasing the accuracy of the results.

Table 3.5 Scaled vs unscaled C and gamma (G) parameters in LIBSVM

	Scaled	Unscaled
Time-Window 1 2014	C: 32 G: 0.5	C: 512 G: 8
Time-Window 2 2014	C: 512 G: 0.5	C: 32768.0 G: 0.5
Time-Window 3 2014	C: 2048.0 G: 0.125	C: 2048.0 G: 2.0
Time-Window 1 2015	C: 1280 G: 0.0078125	C: 32 G: 0.5
Time-Window 2 2015	C: 32 G: 0.0078125	C: 8 G: 0.5
Time-Window 3 2015	C: 128 G: 0.5	C: 2048.0 G: 0.5
Time-Window 1 2016	C: 32 G: 0.0078125	C: 8 G: 0.5
Time-Window 2 2016	C: 32 G: 0.0078125	C: 8 G: 0.5
Time-Window 3 2016	C: 32 G: 0.0078125	C: 32 G: 0.5

DTs were able to produce comparable and even in some cases, higher accuracies than SVMs. One of the reasons for such a high accuracy is because the Kauth-Thomas tasseled cap transformation helps to decrease data dimensionality by only using three bands to express relevant information (Kauth and Thomas, 1976). An increase in data dimensionality decreases the performance of DTs and can cause the Hughes phenomenon, whereby the classifier loses performance as the number of features increases and the training data set stays constant (Mal and Mather, 2003). A reason DTs produced low OA and kappa values from time to time was due to the training sample size. Shao and Lunetta (2012) found that DTs with a low amount of training samples performed poorly compared to SVMs. The same result of using low amounts of training samples occurred in this research when using a conventional classification approach compared to using the time-window method. I acquired training samples for classification from multiple images when using the time-window method, resulting in a higher count of training samples for classification.

How the training samples are distributed in feature space can explain the efficiency or inefficiency of each classification technique. To illustrate this, figures 3.17, 3.18, and 3.19 display the training samples collected from images occurring in time-window 1 for 2014. Each figure shows two of the tasseled cap component values plotted against one another (e.g. Brightness vs Wetness). The splitting of these values in feature space differs for each classification technique. DTs work by splitting the data based on thresholds and because the data in figure 3.17 shows relatively clear separations, DTs perform comparably well against SVMs, a more sophisticated technique. While SVMs provided high OA and kappa values for the study, it was outperformed by DTs on more than one occasion. As with each classification technique, SVMs misclassified the emergent vegetation and open water classes. This misclassification is due to the SVMs inability to optimally separate them in feature space and is a problem noted in other studies that use SVMs (Szantoi et al., 2013). I used the radial basis function as the kernel for the SVM classification and while it has provided high accuracy for numerous studies (Kavzoglu and Colkesen, 2009; Adam et al., 2014), the decision boundary can still incorporate other land cover pixels due to the curve of the decision boundary. Non-water and emergent vegetation pixels occurring close to one another (Figure 3.17), further explaining the difficulty of separating them for classification no matter which technique is used.

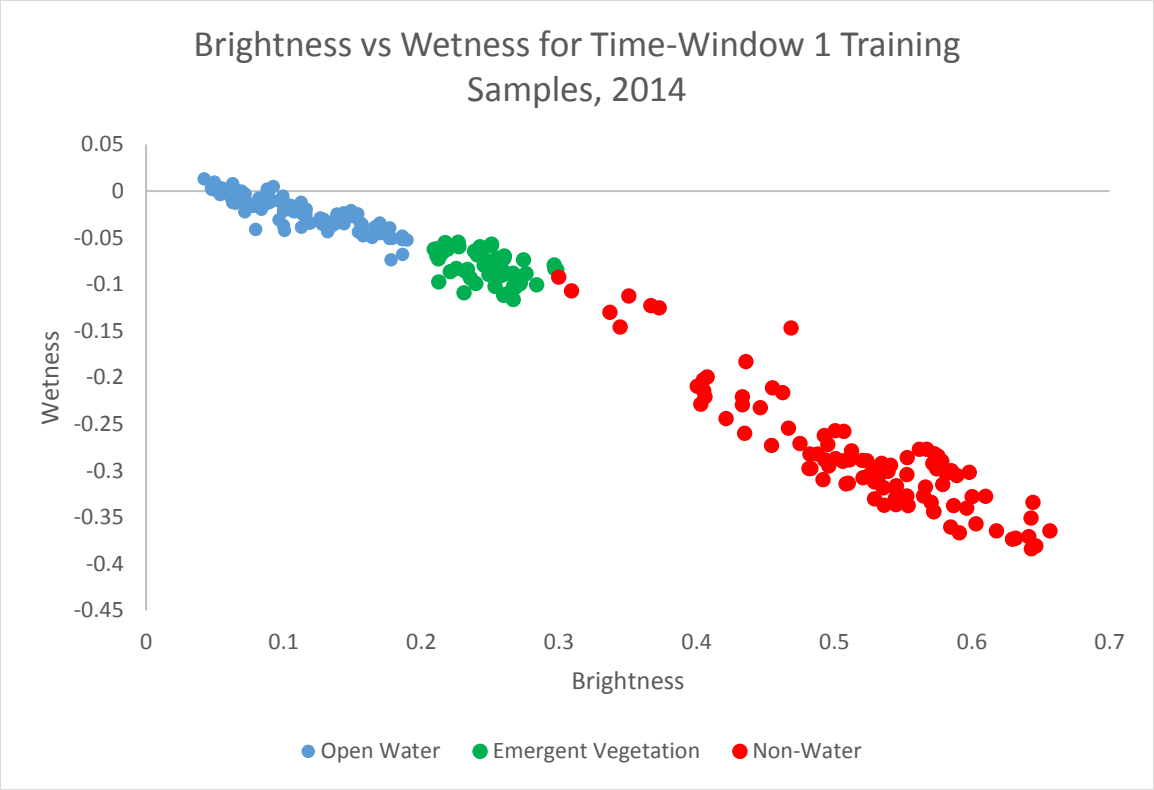


Figure 3.17 Feature space graph for Time-Window 1 land cover training samples with the Brightness and Wetness components

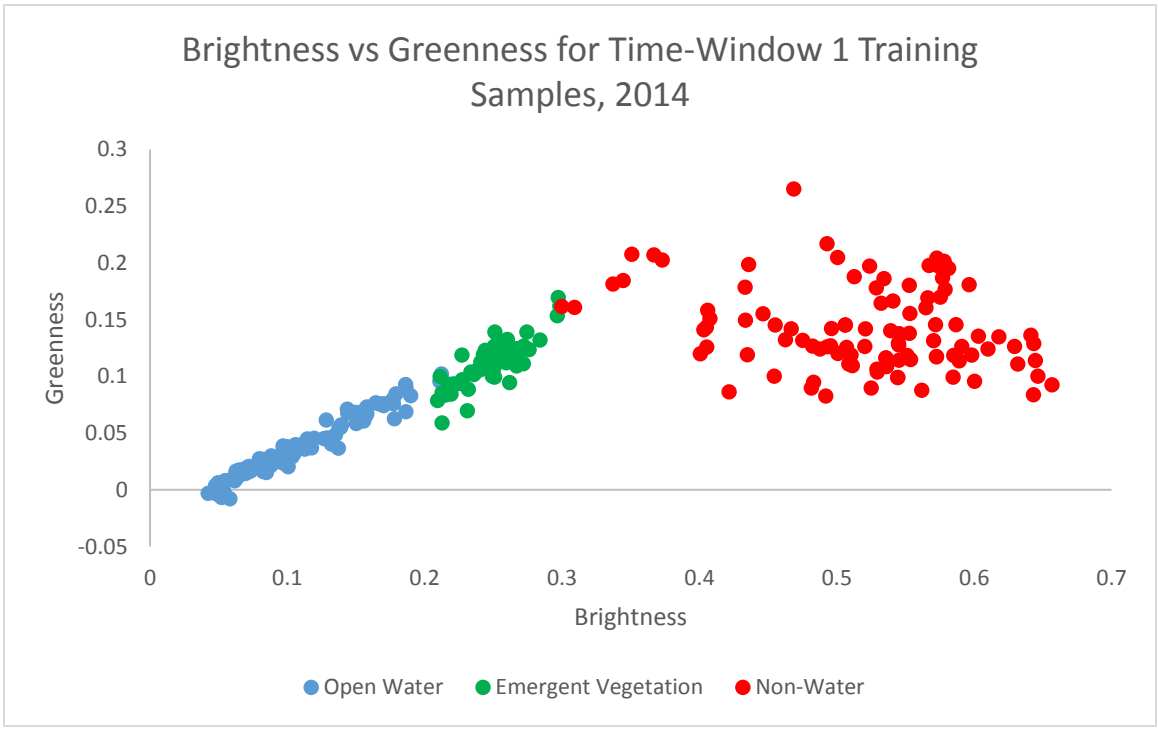


Figure 3.18 Feature space graph for Time-Window 1 land cover training samples with the Brightness and Greenness components

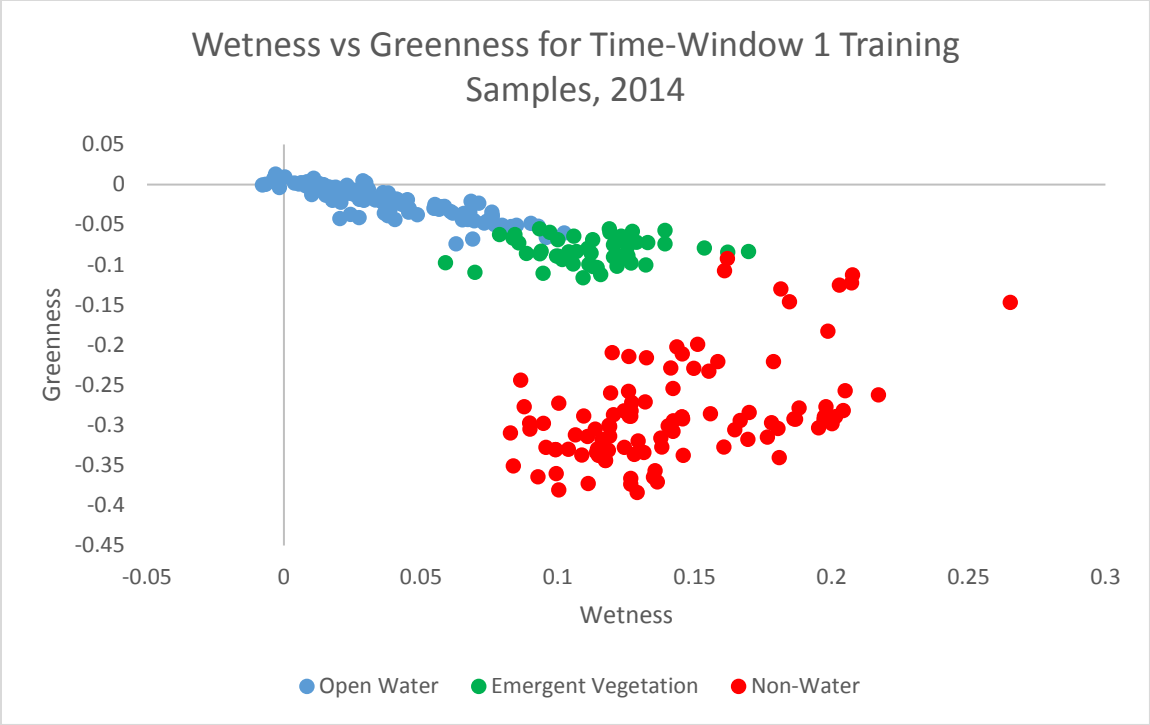


Figure 3.19 Feature space graph for Time-Window 1 land cover training samples with the Wetness and Greenness components

In regards to MLC, the technique has been consistently proven to underperform when compared against SVMs and DTs (Szantoi et al., 2013; Kumar et al., 2014). MLC provided decent OA and kappa values but not adequate enough for the study, nor when compared to SVMs and DTs. Looking again to the feature space graphs (Figures 3.17, 3.18, and 3.19), MLC uses probability functions to define the probability of a given pixel belong to a class. Pixels close to other land covers tend to be misclassified, which can be seen in figure 3.17 where a pixel value is close to the emergent vegetation and non-water classes. Another source of error for MLC is that the algorithm assumes a normal distribution of the training samples. Training samples for the study are not always normally distributed, causing further misclassification (Figure 3.20).

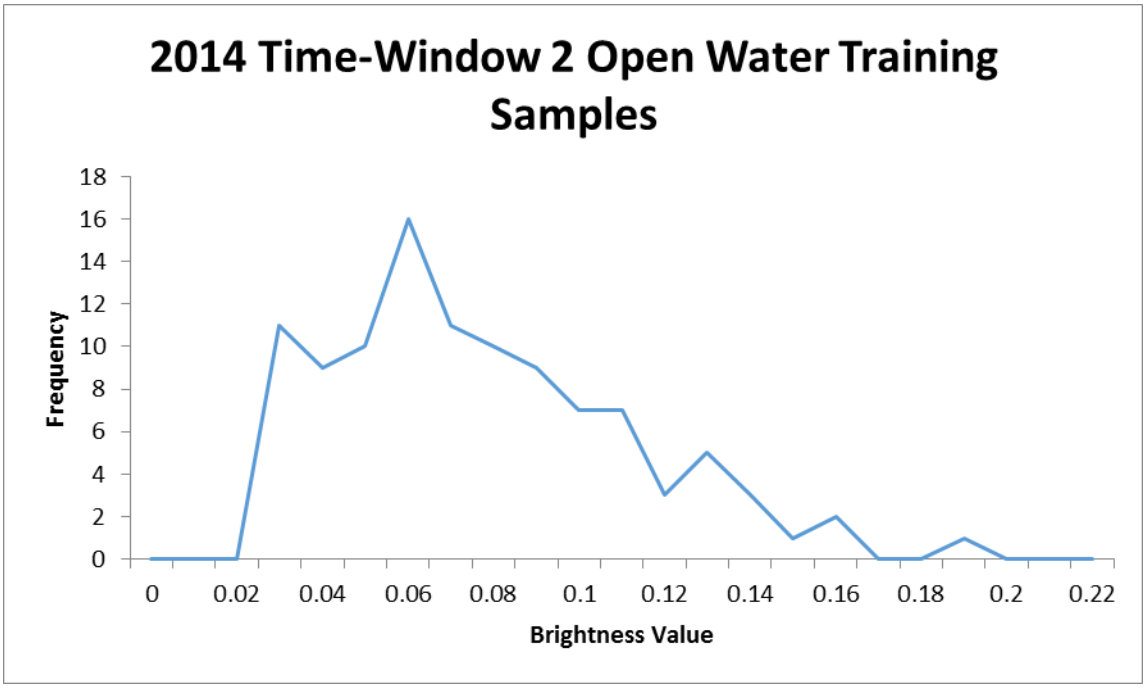


Figure 3.20 Distribution of training samples for the emergent vegetation class

Chapter 4 - Classification Using the Time-Window Method

One of the problems with classifying a series of images through time is that it requires a large amount training sample collection and classification. The amount of time required to provide accurate results is too substantial for most research and results are needed sooner rather than later to superiors. Because of the challenges inherent in classifying a series of images overs time, I present a new classification framework, built around time-windows and a training library. This chapter focuses on addressing the second research objective of creating a framework for classifying imagery using time-windows to speed the classification procedure up without sacrificing accuracy. The chapter covers defining the time-window method and the results of using the method to classify imagery in the CRB. A discussion follows, which focuses on the effectiveness of the time-window method.

4.1 Methods

4.1.1 Defining Time-Windows

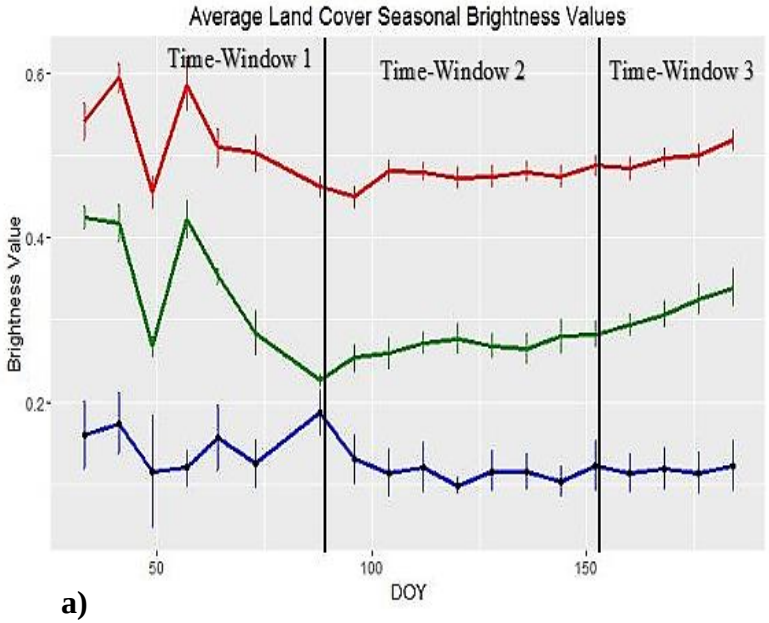
To illustrate both the problem and the approach of classifying flooding extent in the Chobe River basin, consider the temporal trajectories of the Kauth-Thomas tasseled cap component values (Figures 4.1a-c). Recall from chapter 3, figure 4.1a-c are shown again for clarity. I am unsure why there is a v shape to the brightness component values around DOY 50 except the image for that data was darker than the images the occurred before and after that date for the three years. From day of year (DOY) 33 to roughly DOY 91, the land covers of emergent vegetation and non-water exhibit decreasing values, while open water values remain consistent. Starting around DOY 91 to roughly DOY 155, all three land covers exhibit relatively constant values or show very slight changes in values from composite to composite. After DOY 155, both emergent vegetation and non-water show increasing values, while open water values remain consistent. Using the same training data set to classify images over the entire study period is therefore clearly not feasible due to these changes in values throughout the flooding period. However, the temporal variation of these trajectories suggests that they can be divided into three subperiods of roughly coherent behavior, and that training samples collected during these periods may therefore be representative of, and applicable to, all of the images within them. Here, these subperiods will be referred to as time-windows 1, 2, and 3. The time-windows help to group

similar biophysical property values and maintain the accuracy of a traditional supervised classification image.

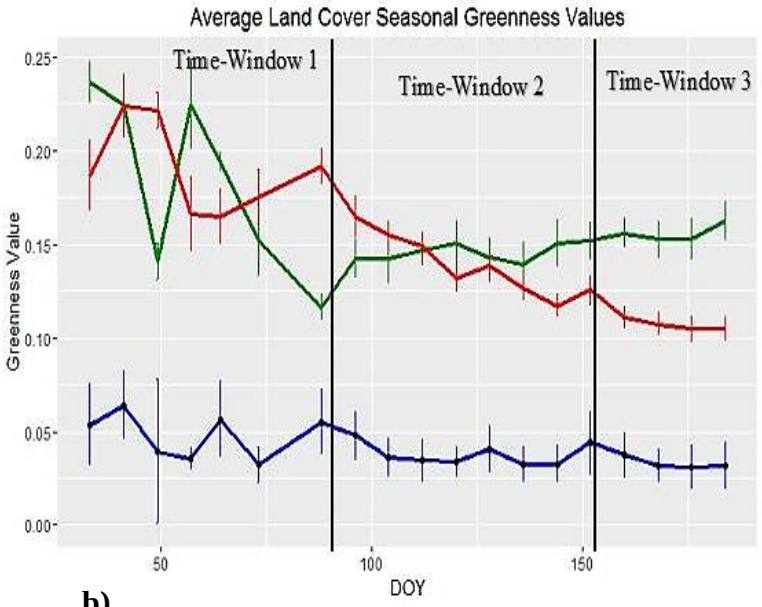
To provide more details, figure 4.1a shows that from February to March, the values of emergent vegetation change from 0.43 to 0.23. This change in values requires the study period to be separated further into three time-windows. Each time window was separated based on biophysical properties of each land cover type no longer showing similar values associated with each other. For example, time-window 1 consists of emergent vegetation values decreasing until March 29. Time-window 2 has consistent values for emergent vegetation from the beginning of April until early June. Time-window 3 shows the senescence of the emergent vegetation from early June to July. The three time-windows specifically are; February to the end of March (time-window 1), beginning of April to the beginning of June (time-window 2), and from June to July (time-window 3). I used the same threshold dates across all three years of the study period to delineate the time-windows. The reason for using the same dates is because after collecting the land cover trajectories of a given year and then averaging them out across the study period, the results showed that they had similar trajectories. The low standard error associated with the threshold dates shows that around the same time each year, the component values start to disassociate from previous dates and start to exhibit different component values. In other words, the land cover trajectories do not change that much year to year and therefore using the same dates is not going to impact the accuracy results.

To operationalize this approach, a set of Kauth-Thomas component values for each cover type (see chapter 3.1 for discussion of Kauth-Thomas tasseled cap), from each of the three time-windows were collected by 'head up' selection of pixels from locations known to represent each cover type. These grouped pixel values were used to form a library of 'signatures' representing each of the cover types, through time, which were then used to classify the imagery. This library of signatures is defined here as a training library. As an example, for time-window 1 of a given year, four to five images need to be classified. Training samples were collected from two to three images in each time-window. For each of the selected images, 50 – 60 training sites were chosen per land cover. These samples were used to create a training library, which were used to train each decision rule. I used the same three decision rules in chapter 3 to classify the imagery. It is important to note that images were classified using training samples collected from the time-window they occur in and from that year. As an example, training samples collected in time-

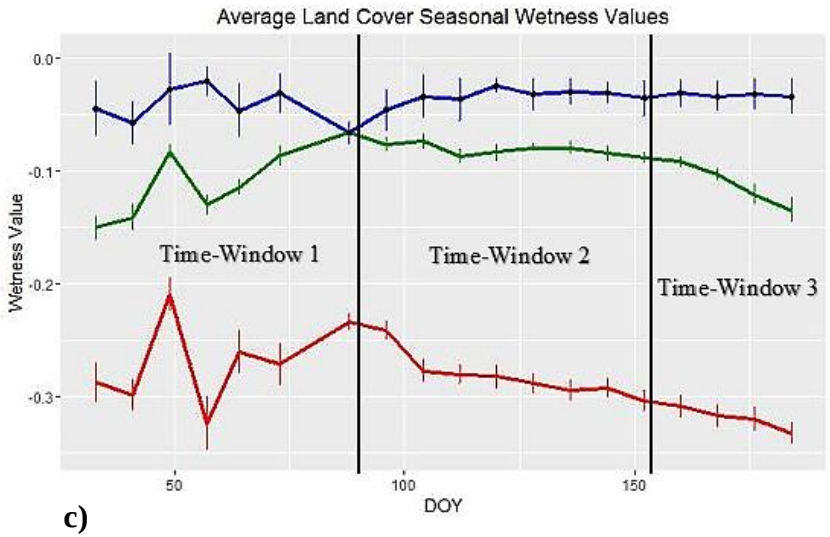
window 1 for 2014 were used to classify imagery during February and March of 2014. In other words, I developed one training library for each time-window for a total of three in one year. A different training library was developed for time-window 1 for 2014 compared to time-window 2 for 2014. Grouping images with similar Kauth-Thomas component values in this framework of time-windows and using a library of signatures will maintain accuracy and decrease the time consumption necessary to classify the images for this analysis.



a)



b)



c)

Figure 4.1 a-c Seasonal tasseled cap component values for open water (blue), emergent vegetation (green), and non-water (red), from 2014 to 2016. Figures 4.1a, b, and c display the brightness greenness, and wetness component, respectively, land cover values through the study period.

4.2 Results of Time-Window Classification

The following three subsections detail the overall accuracy and kappa values for each time-window to determine if splitting the study period into these three time-window is an effective framework for classification. See appendix A for a full list of error matrices made for each image validation. All three decision rules; SVMs, DTs, and MLC, were used for classifying imagery that occurs in each of the time-windows. While it was determined that DTs and SVMs provided the highest performance for the CRB, each decision rule was used for classification. I acknowledge that there is some circular logic to these methods.

4.2.1 Classification Results for Time-Window 1

For the years 2014 and 2016, clouds affected the accuracy of some of the classifications for time-window 1 and resulted in low overall accuracy and kappa values. The year 2015 had less cloud cover, which resulted in better overall accuracy (OA) and kappa values. For time-window 1, one evaluation image was chosen to represent the accuracy for the time-window for a given year. The dates of these images are: March 14, 2014, 2015, and February 10, 2016. For example, a total of three images were created for accuracy assessment for time-window 1 in 2015, because all three decision rules were run on the image. Each evaluation image was selected based on the availability of a Landsat 8 reference image and minimal clouds occurring in the study area.

DTs performance was higher than SVMs and MLC for the study period of March 14, 2014 (table 4.1). SVMs had a kappa of 0.54, which was surprising because SVMs are considered a higher level decision rule. For the study March 14, 2015 study period, the performance of SVMs and DTs were virtually the same. Both SVMs and DTs had high kappa values of 0.84, indicating substantially better performance than random chance. SVMs provided the highest performance for the February 10, 2016 study period. I calculated the average OA and kappa values across all three years for time-window 1 and found that SVMs and DTs were virtually the same at mapping the land covers.

Table 4.1 Average OA and Kappa for time-window 1 for all three classification techniques across the study period: 2014 – 2016

	March 14 2014	March 14 2015	Feb 10 2016	2014-2016
DT OA	0.8	0.9	0.76	0.82
DT Kappa	0.7	0.84	0.63	0.72
SVM OA	0.7	0.9	0.85	0.82
SVM Kappa	0.54	0.84	0.78	0.72
MLC OA	0.64	0.85	0.76	0.75
MLC Kappa	0.48	0.78	0.62	0.62

All classification technqiues OA - 0.80, kappa - 0.69

4.2.2 Classification Results for Time-Window 2

For time-window 2, all the imagery was cloud free, resulting in both high OA and kappa values for all three classification techniques. The same process of accessing the accuracy of the time-window stated in section 4.1.1 was followed with the image dates of: May 17, 2014, May 25, 2015, and May 23, 2016. For the May 17, 2014 study period, the performance of the DT and SVM were virtually the same (table 4.2). For the May 25, 2015 study period, the performance of SVMs was higher than that of DTs and MLC, with MLC outperforming DTs. SVMs provided the highest performance for the May 24, 2016 study period, but the overall performance of SVMs, DTs, and MLC were virtually the same. Looking across all three years for time-window 2 and found that SVMs provided the highest performance, and DTs and MLC produced similar results.

Table 4.2 OA and Kappa values for each classification technique in time-window 2

	May 17 2014	May 25 2015	May 24 2016	2014-2016
DT OA	0.89	0.86	0.9	0.88
DT Kappa	0.83	0.79	0.85	0.82
SVM OA	0.88	0.94	0.93	0.92
SVM Kappa	0.82	0.9	0.9	0.87
MLC OA	0.87	0.89	0.87	0.88
MLC Kappa	0.81	0.84	0.81	0.82

All classification techniques OA - 0.89, kappa - 0.84

4.2.3 Classification Results for Time-Window 3

During time-window 3, open water coverage decreased, and dry land and emergent vegetation cover increase. The same process of accessing the accuracy of the time-window stated in section 4.1.1 was followed with the image dates of: June 18, 2014, June 25, 2015, and June 25, 2016. SVMs provided the highest performance for the June 18, 2014 study period, while the performance of DTs and MLC were virtually the same (table 4.3). For the June 26, 2015 study period, DTs provided the highest performance, followed by SVMs. For the June 25, 2015 study period, the performance of all three decision rules were virtually the same, with SVMs providing the lowest OA and kappa values. Across all three years of the study period, DTs and SVMs provided the highest performance for time-window 3.

Table 4.3 OA and Kappa values for each classification technique in time-window 3

	June 18 2014	June 26 2015	June 25 2016	2014-2016
DT OA	0.9	0.95	0.91	0.92
DT Kappa	0.85	0.92	0.87	0.88
SVM OA	0.95	0.9	0.87	0.91
SVM Kappa	0.91	0.85	0.8	0.86
MLC OA	0.9	0.84	0.9	0.88
MLC Kappa	0.85	0.76	0.86	0.82

All classification techniques OA - 0.90, kappa - 0.85

There were slight differences in kappa but overall, SVMs and DTs were equally effective for classification of imagery that occurred in their respective time-windows 1, 2, and 3. Time-window 1 suffered from cloud cover, resulting in low OA and kappa values. Meanwhile, time-windows 2 and 3 had high OA and kappa values for each of the decision rules tested.

4.2.4 Comparison of Time-Window Method against Other Methods of Classification

To further address the second question of the study, I used three methods to determine the effectiveness of the time-window method and they are broken down here. Method I involves classifying imagery with training samples collected in the time-window the image occurs in

using the time-window method. Method II involves classifying imagery with training samples collected from imagery that occurs in time-window 2. Method III is a conventional supervised classification, where training samples are collected from an image and are used to classify that image. A brief section detailing Methods II and III follow below. These methods are evaluated against one another using the z-statistic to report a p value, which is calculated based on the variance of the kappa values for the two methods being compared.

Method II

I classified imagery that occurred in time-windows 1 and 3 with training samples collected from time-window 2 to demonstrate the effectiveness of splitting up the study period into three time-windows as opposed to using the same training samples for the entire year. The image for time-window 1 was March 14, 2015 and was compared against the March 14, 2015 image classified using Method I, or the time-window method. The procedure was done for the time-window 3 image, June 26, 2015. Method II underperforms based on OA and kappa values compared to Method I (table 4.4). As an example, the DT result for June 26, 2015 (Figure 4.2) is compared to its counterpart using samples collected from time-window 2 (Method II) (Figure 4.3). Figure 4.3 tends to under predict the open water class. The reason for the lower OA and kappa values is a result of the changing tasseled cap component values through the study period (Figure 4.1). For instance, MLC classifies emergent vegetation as water because emergent vegetation values in time-window 2 are close to water values in feature space, therefore the probabilities think the emergent vegetation values are water.

Table 4.4 OA and Kappa values for Methods I, II, and III for time-windows 1 and 3

	Time-Window 1			Time-Window 3		
	Method I	Method II	Method III	Method I	Method I	Method II
DT OA	0.9	0.79	0.87	0.89	0.95	0.82
DT Kappa	0.84	0.69	0.80	0.83	0.92	0.73
SVM OA	0.9	0.8	0.88	0.90	0.9	0.83
SVM Kappa	0.84	0.71	0.82	0.85	0.85	0.74
ML OA	0.85	0.79	0.87	0.87	0.84	0.83
ML Kappa	0.78	0.68	0.80	0.81	0.76	0.75

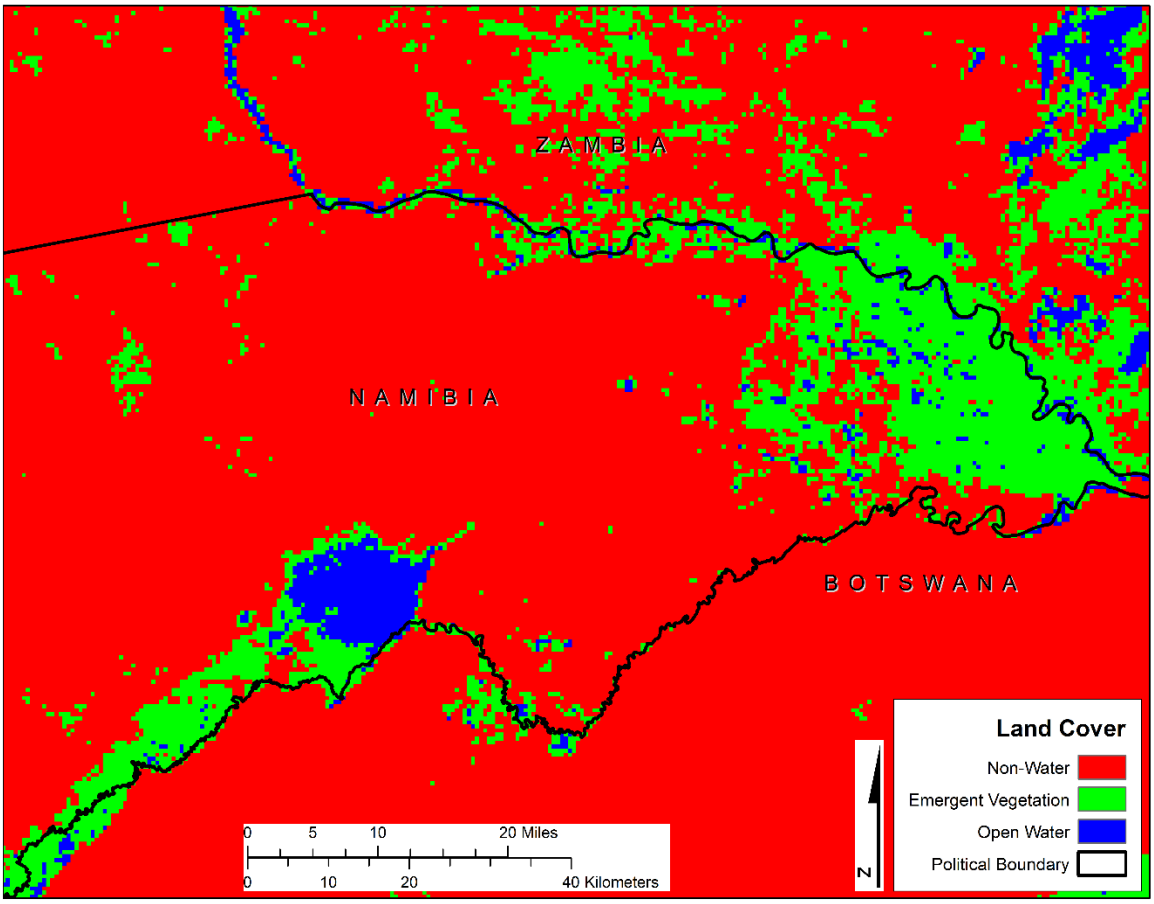


Figure 4.2 June 26, 2015 classified imagery using DT

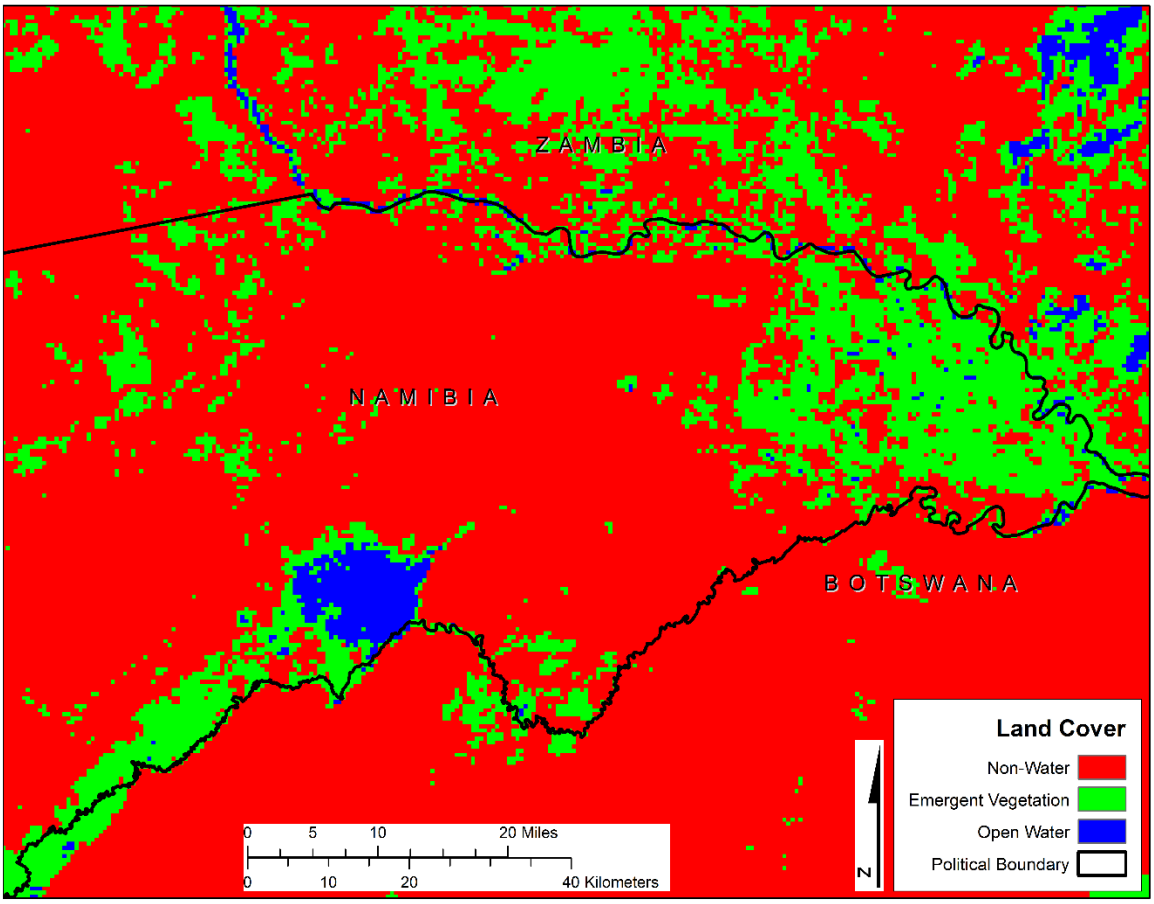


Figure 4.3 June 26, 2015 classified imagery using DT and using time-window 2 training samples

Method III

The conventional supervised classification approach was done to compare against the classification results using the time-window method. 50 to 60 training sites were chosen for the three land covers for each of the 3 images in 2015; March 14, May 25, and June 26. One image occurs in each of the time-windows to get an understanding of the effectiveness of classification of imagery that occurs in each of the time-windows. 2015 was chosen because of MODIS data availability and had the highest accuracy for time-window 1 compared to 2014 and 2016.

As expected, the conventional classification technique proved to be successful. All techniques show an acceptable OA and kappa for classification for the region except DT on May 25, 2015 (table 4.5). DT performed poorly in terms of OA and kappa values with the limited amount of training samples for this one image, further reinforcing the literature in that DTs do not do well with a limited amount of training samples (Pal and Mather, 2003).

Table 4.5 OA and kappa values for the conventional supervised classification

	March 14 2014	May 25 2015	June 26 2015
DT OA	0.87	0.76	0.89
DT Kappa	0.80	0.64	0.83
SVM OA	0.88	0.87	0.90
SVM Kappa	0.82	0.81	0.85
MLC OA	0.87	0.88	0.87
MLC Kappa	0.80	0.83	0.81

4.2.5 Comparison of Time-Window Method Using Z-Score

The z-score (p-value) (table 4.6) and the kappa (variance) (table 4.7) are reported for the comparison of Method I against Method II and Method III. For example, the z-score between Method I and Method II for March 14, 2015 using DT was 1.376 and a p-value of 0.9147. I found that Method I was significantly better than Method II (z-1.67, p-0.9525) for June 25, 2015 using the DT decision rule. The significant value here effectively means that splitting the study period into three time-windows is an effective means to classify imagery in the CRB instead of using the same training samples across the entire study period. When comparing the remaining images of the kappa variances between Methods I and II, there were no more statistically significant p-values found, but when comparing the classified imagery with the reference data, the imagery classified with Method I represent the reference data more effectively. When comparing Method I and III, there were no statistically significant p-values found, which is ideal because it demonstrates that the time-window method of classification (Method I) is effectively the same as a conventional supervised classification technique (Method III). Using the time-window method cuts down on classifying each image and ultimately saves time without sacrificing accuracy of mapping land cover in the CRB.

Table 4.6 Comparison of Method I against Methods II and III for each decision rule using a z-score (p-value)

Date	MII	MIII
March 14, 2015 - SVM	1.20(0.8849)	0.183(0.5714)
March 14, 2015 - DT	1.376(0.9147)	0.367(0.6406)
March 14, 2015 - ML	0.091(0.5359)	-0.181(0.4286)
May 26, 2015 - SVM	0.806(0.7881)	0.806(0.7881)
May 26, 2015 - DT	1.277(0.8980)	1.277(0.8980)
May 26, 2015 - ML	0.186(0.5714)	0.186(0.5714)
June 26, 2015 - SVM	0.981(0.8365)	0 (0.5)
June 26, 2015 - DT	1.67(0.9525)	0.787(0.7823)
June 26, 2015 - ML	0.091(0.5359)	-0.449(0.3300)

Table 4.7 Kappa(variance) for each of the methods tested to determine the effectiveness of the time-window method

Date	MI	MII	MII
March 14, 2015 - SVM	0.84(0.008)	0.71(0.005)	0.81(0.005)
March 14, 2015 - DT	0.84(0.007)	0.69(0.005)	0.80(0.005)
March 14, 2015 - ML	0.78(0.007)	0.68(0.005)	0.80(0.005)
May 26, 2015 - SVM	0.90(0.007)	0.81(0.005)	0.81(0.005)
May 26, 2015 - DT	0.84(0.007)	0.82(0.005)	0.82(0.005)
May 26, 2015 - ML	0.79(0.008)	0.64(0.006)	0.64(0.006)
June 26, 2015 - SVM	0.85(0.007)	0.74(0.005)	0.85(0.005)
June 26, 2015 - DT	0.92(0.008)	0.73(0.005)	0.83(0.005)
June 26, 2015 - ML	0.76(0.007)	0.75(0.005)	0.81(0.005)

4.3 Discussion of Time-Window Method

The results submitted here suggest that using the time-window techniques attains classification accuracy on par with a conventional classification technique. Also, based on the p-values, the time-window technique (Method I) was significantly better than using training samples from other time-windows (Method II) when the comparison was made for one image, June 25, 2015 using DT. For the rest of the dates that were considered, p-values were close to the threshold of significance but no other significant values were found. Using time-windows for a classification framework provides results that are similar if not better than using training values from a different time-window. The trajectories of the tasseled cap component values for each of the land covers in the CRB are the reason for the success of the method. As noted previously, the land covers are not constant through the study period and using values from one part of the study period (February) to classify land cover in another part of the study period (May), results in low accuracy results. Collecting training samples that exhibit coherent behavior across dates for a single time-window does not confuse the decision rules during the classification process.

No significant p-values were found between Methods I and III and this is ideal because it shows that classifying imagery with the time-window method is just as effective as classifying imagery using a conventional supervised classification procedure. Classifying imagery with training samples collected from the same image will usually result in high accuracy results depending on the decision rule. It is not surprising that a conventional supervised classification performed well, but occasionally the time-window method provided higher accuracy results. Collecting training samples from multiple images captures more variance of the land covers in the study area. The time-window method also incorporates more training samples for the decision rule to base their decisions on, compared to the conventional supervised classification that used 50-60 training samples for each land cover. The lower amount of training samples resulted in the DT providing low accuracy results compared to the time-window method for the same date for one image.

Chapter 5 - Flooding Extent in the CRB

The third objective seeks to determine if the flooding extent in the CRB was decreasing or increasing from 2014 to 2016 and compare the flooding extent totals to past work focusing on the CRB. This chapter briefly explains how I quantified flooding in the CRB and reports the results of the inter and intra-annual flooding extents. A brief discussion is presented on reasons for increasing or decreasing flooding in the CRB.

5.1 Methods

The flooding extent was calculated using the DT and SVM images because they provided the best accuracy results. Both were used to quantify the flooding extent to see if there were major differences in how they quantified the flooding extent between the emergent vegetation and open water classes. I summed the number of pixels in each image that belonged to the emergent vegetation and open water classes, and multiplied the total by 250,000, the area of one MODIS pixel, before converting the values to km^2. It should be noted that flooding extent was quantified in the Zambezi Wetlands, Chobe River, and Lake Liambezi areas, and not the Mamili Wetlands.

5.2 Results of Flooding Extent

DT had a tendency to report more of the emergent vegetation class, while SVMs reported more of the open water class. Using SVMs (Figures 5.1 – 5.4), the maximum flooding extent for 2014 and 2015 was June 10 at 3603.25 km^2 and 2584.5 km^2, respectively. The maximum flooding extent for 2016 occurred on April 30 at 2564 km^2. SVMs had a tendency to classify more of the emergent vegetation class than the open water class. Using DTs (Figures 5.5 – 5.8), the maximum flooding extent for 2014 and 2015 was also June 10 at 3698.75 km^2 and 2225.5 km^2. The maximum flooding extent for 2016 was slightly less than SVMs at 2558 km^2. DTs classified more of the land cover in the region as the open water class than the emergent vegetation class. 2014 saw the largest extent (Figure 5.4, 5.8). 2015 was a significantly reduced flooding event but SVMs classified more of the two flooding classes, 2584.5 km^2, than DTs, 2225.5 km^2. Looking at 2015 and 2016, SVMs show comparative flooding extents, while DTs show a smaller flooding event in 2015 and then an increase in 2016. The differences in flooding

extent for the 2015 season changes the perception of flooding even though in 2016, both SVMs and DTs classified comparative flooding extents, 2564 km^2 and 2558 km^2.

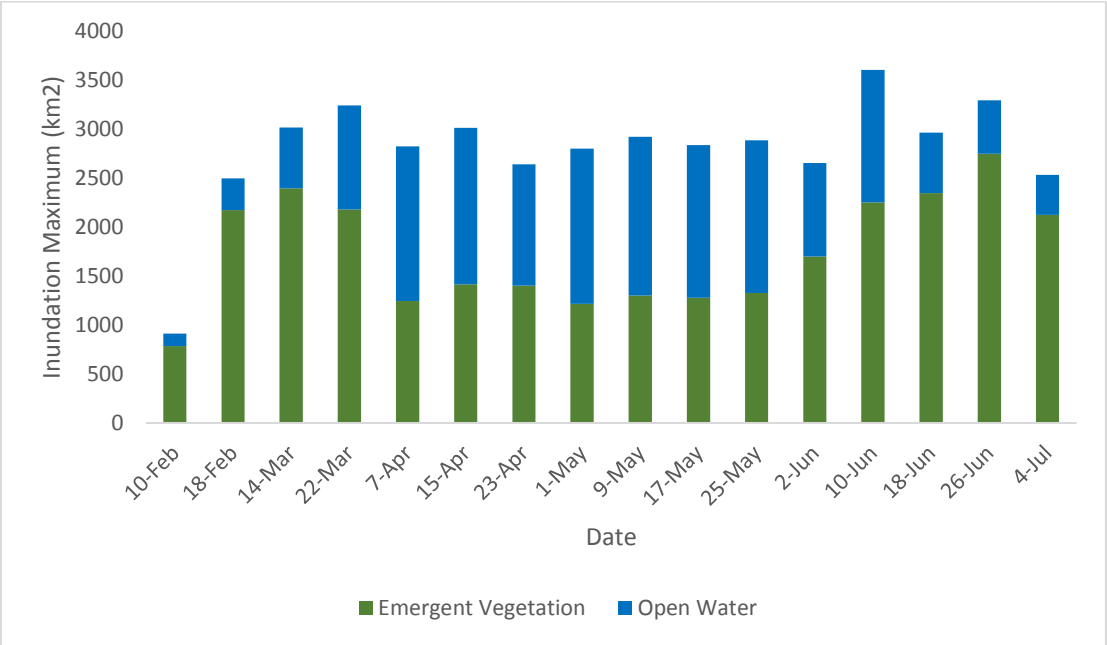


Figure 5.1 2014 seasonal flooding extent using SVM

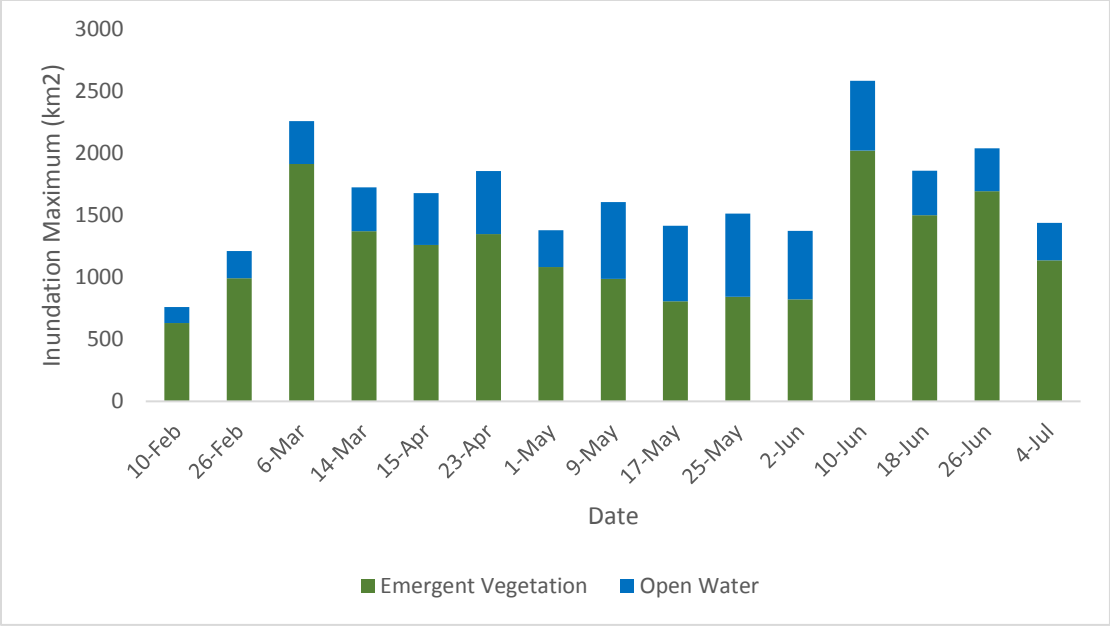


Figure 5.2 2015 seasonal flooding extent using SVM

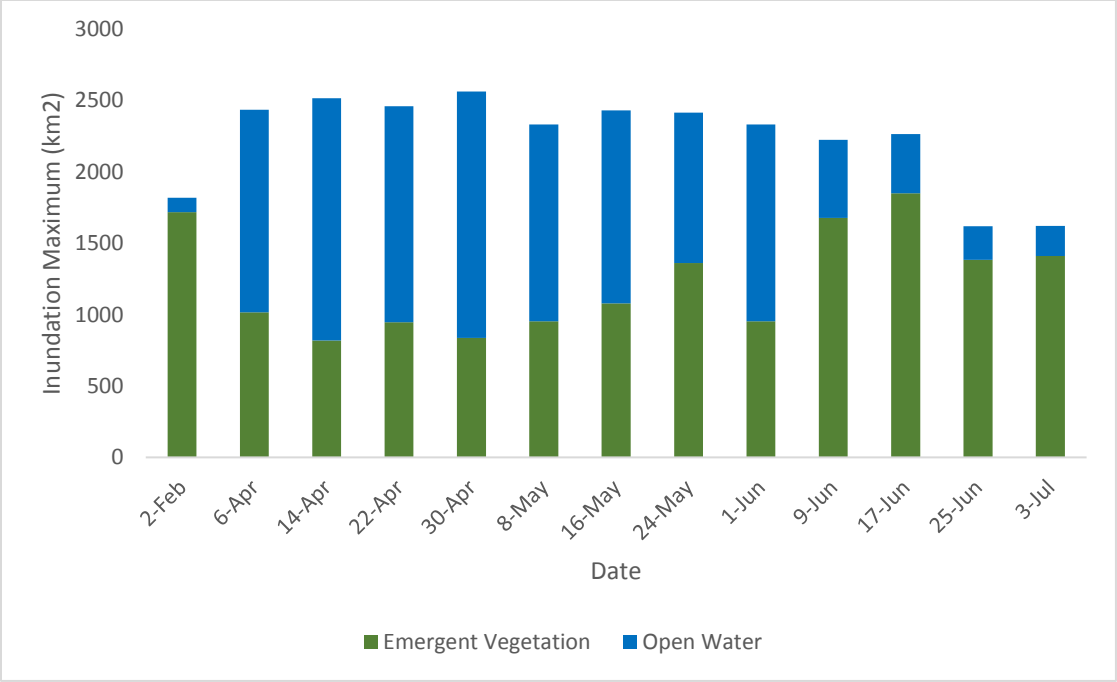


Figure 5.3 2016 seasonal flooding extent using SVM

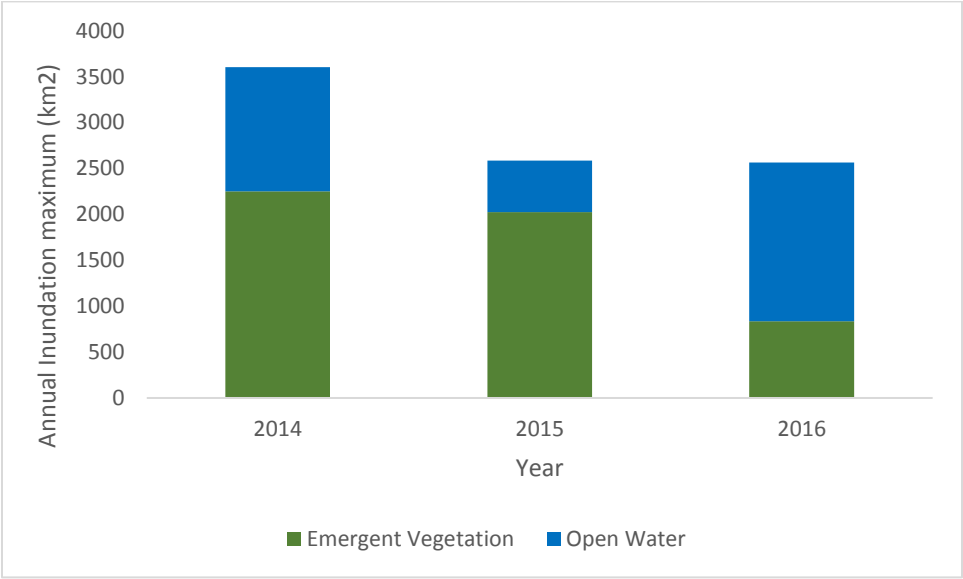


Figure 5.4 Annual inundation maximums from 2014 to 2016 using SVM

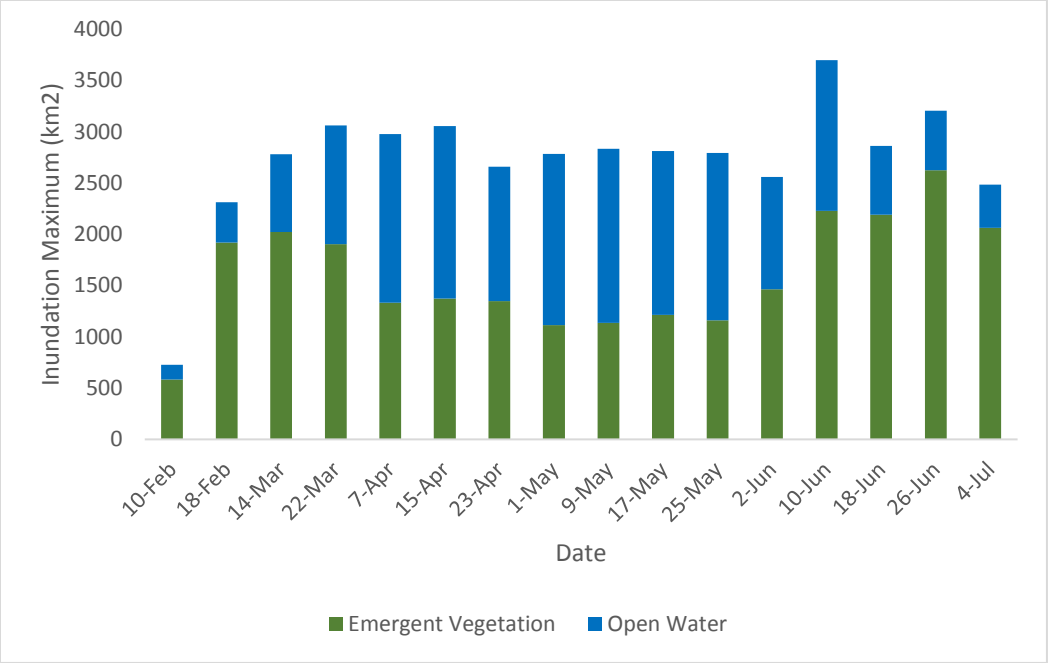


Figure 5.5 2014 seasonal flooding extent using DT

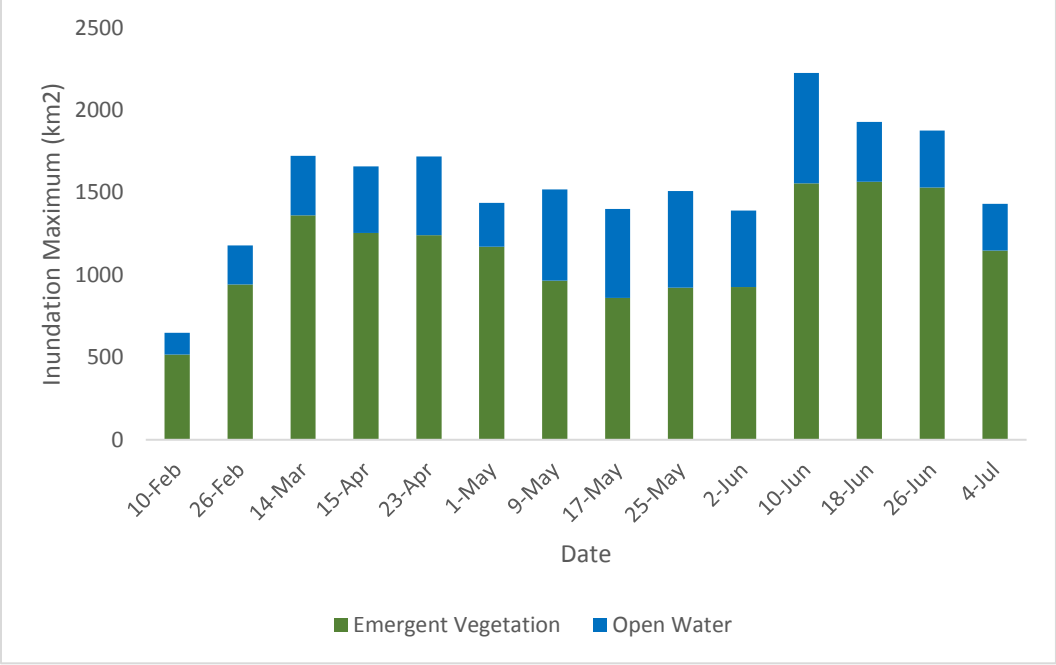


Figure 5.6 2015 seasonal flooding extent using DT

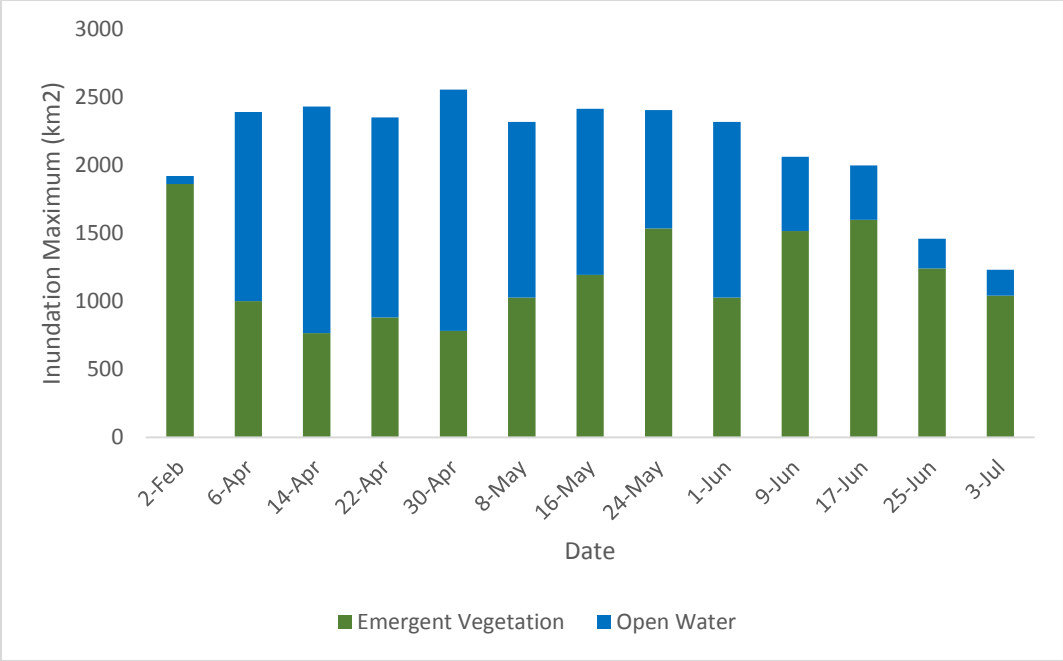


Figure 5.7 2016 seasonal flooding extent using DT

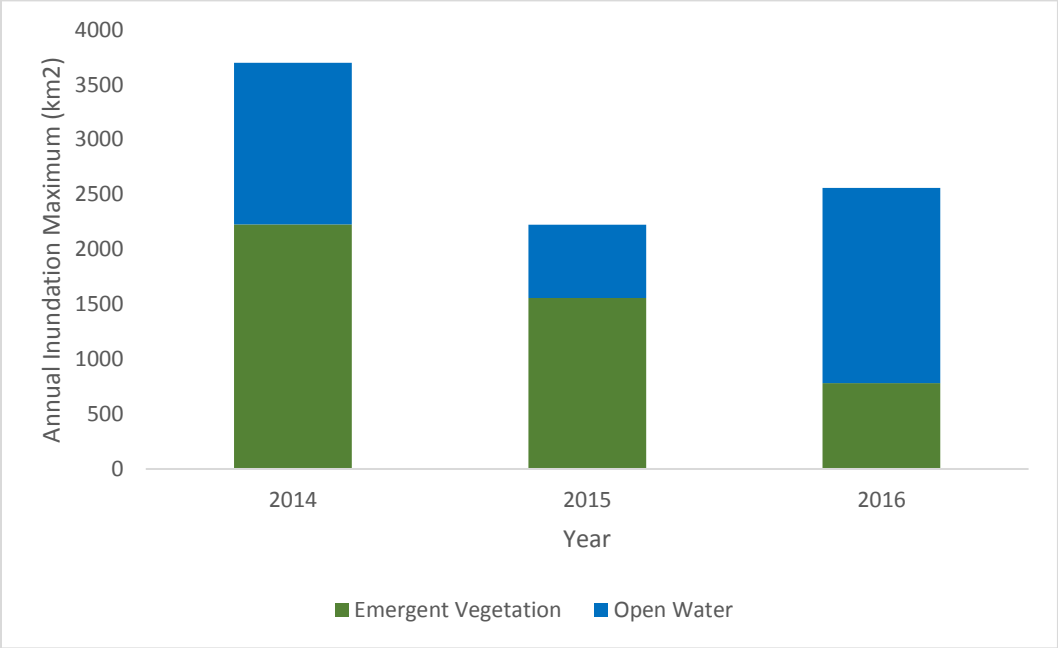


Figure 5.8 Annual inundation maximum from 2014 to 2016 using DT

5.3 Discussion of Flooding Extent

The study analyzed three years of flooding extent in the Chobe River Basin. As previously mentioned, our third objective was to map the extent of flooding in the CRB and compare the data to studies done by Pricope (2013) and Burke et al. (2016) to determine if the trend of flooding is increasing or decreasing. For both 2014 and 2015, the maximum flooding extent occurred on June 10. Studies by Burke et al (2016) and Pricope (2013) support these findings because they found the peak flooding extent to be mid-June. The maximum flooding event occurs in mid-June because the Kwando River flood pulse in May-June arrives to further inundate the area (Burke et al., 2016). The peak maximum flooding extent for 2016 occurred on April 30 and then remained slightly below that peak until tapering off at the end of June. The smaller flooding extent compared to 2014 is because the secondary flood pulse for 2016 only contributed to Lake Liambezi and because the lake was at a low level, only small amounts of water reached the Chobe River. Also, the second flood pulse for 2016 was relatively small judging from the imagery, contributing only small amounts of water to the CRB. Flooding in the CRB is also a building process from multiple flood pulses and as a result of the small 2015 flood (Figure 5.9), less water was available to remain in Lake Liambezi, thereby decreasing the maximum flooding extent the next year (Figure 5.10).

The analysis of annual flooding maximums for the study period of 2014 to 2016 visually reveals a decreasing overall trend. Burke et al. (2016) noted a positive overall trend of flooding extent since 2000, with the high magnitude floods experienced in the CRB since 2009 to 2014 as the main cause for the positive trend. A reason for a possible decreasing trend as noted by Burke et al. (2016), is that warm phase ENSO causes low spatial extent of flooding, as it did in 2002 and 2005. 2015 experienced an El Nino, and would explain the low spatial extent of flooding for that year. However, Burke et al. (2016) also mention that the El Nino event of 2009/2010 brought above average rainfall conditions to the region. It is tough to make a conclusion as to why the flooding extent in the region is moving towards a decrease. One possible explanation is that the CRB is starting to experience a dry cycle. Looking at archival MODIS imagery from as far back as 2000, the spatial extent of flooding in the CRB seems to be cyclical, possibly around 10 years. This cycle is corroborated by the study done by Burke et al. (2016), which found small flooding extent around the year 2005. However, a longer annual flooding record is needed to make a more educated assertion if there is an increasing or decreasing trend.

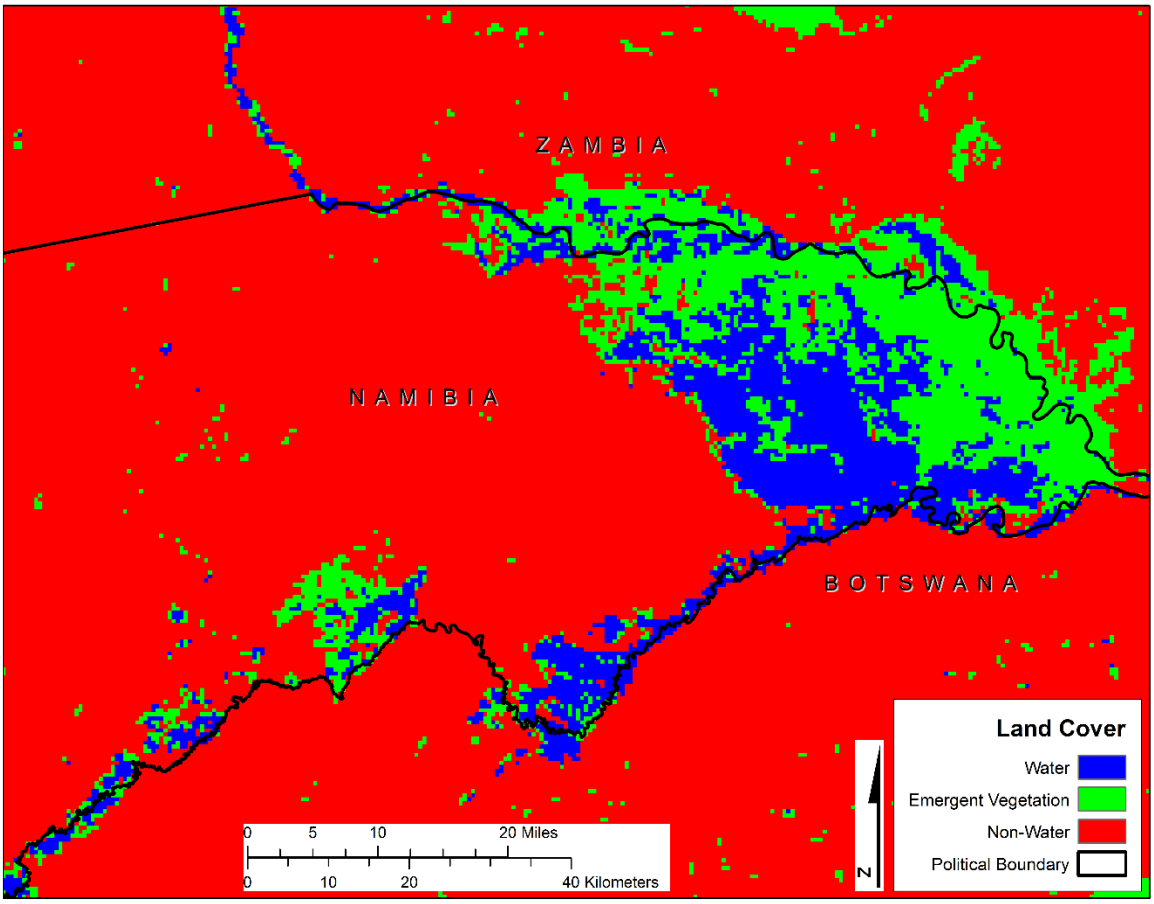


Figure 5.9 May 24, 2016 classified imagery using SVM

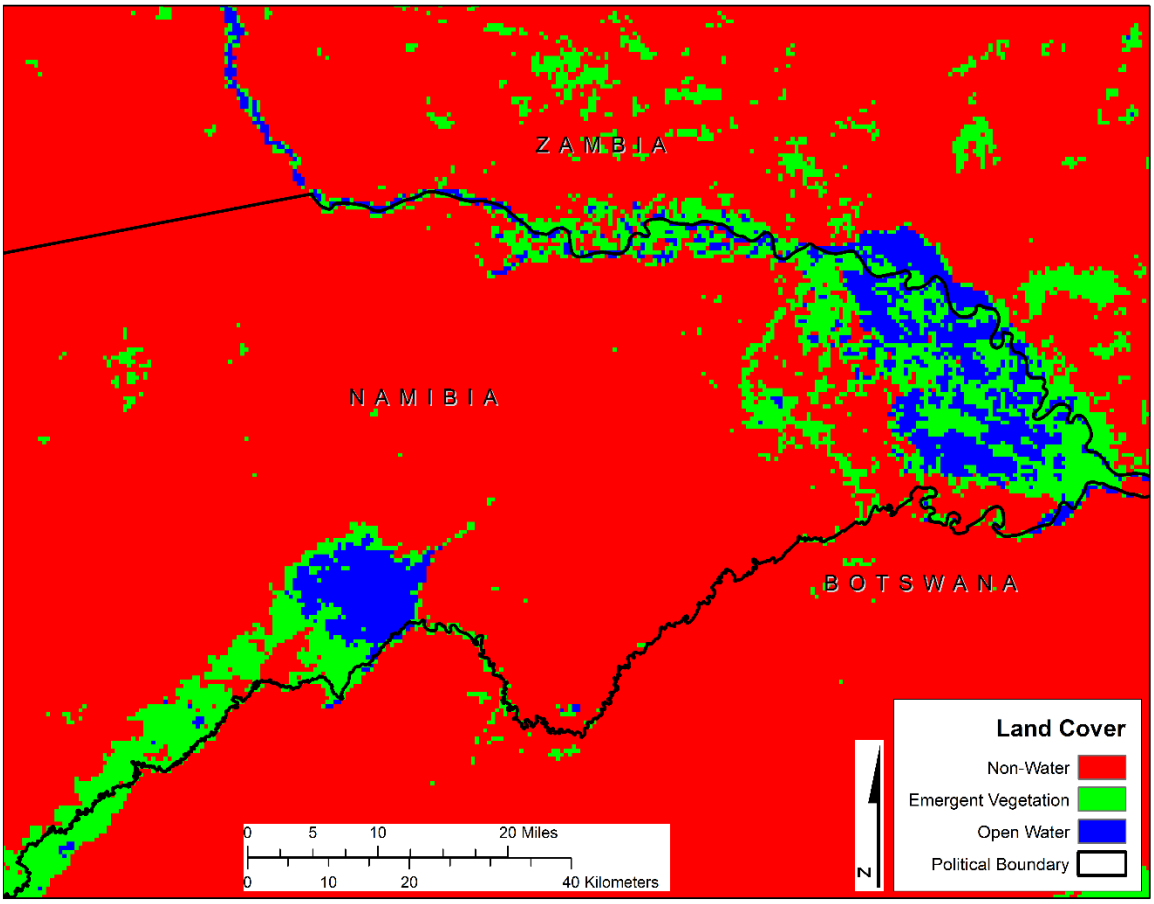


Figure 5.10 May 9, 2015 classified imagery using DT

Chapter 6 - Conclusion

A new technique for classifying multiple images over a study period was proposed. The technique takes into account the changing tasseled cap values for each of the three land covers in the study over the entire study period. I split the study period into three time-window to help with classification. Each time-window was grouped based on land covers exhibiting roughly coherent behavior. Training samples were assembled into a training library that classified imagery that occurred in a specific time-window. The time-window method proved to provide high accuracy results, even comparable to a traditional classification approach. When using training samples collected from another time-window to classify imagery, the accuracy results decreased. Using a z-statistic showed some significance between the kappa values and a visual inspection revealed a decrease in quality of representing the land covers. The time-window method therefore helps to cut down on classifying each image and ultimately save time without sacrificing accuracy of mapping land cover in the CRB.

SVMs produced the best accuracy results for time-window 2, while DTs produced the best results for time-windows 1 and 3. SVMs had the highest overall accuracy and kappa values for the entire study period (2014 – 2016), with DTs producing comparative results. SVMs were also the best classifier for 2015 and 2016. DTs produced the best results in 2014. There were slight differences in kappa but overall SVMs and DTs were equally effective, and it is therefore difficult to recommend one over the other. However, I believe DTs produce the best results because they have similar OA and kappa values to SVMs but they do not over predict the 2015 flooding extent like SVMs do. As expected, MLC had the lowest accuracy results, which can be attributed to the data not being normally distributed. The major source of error for all classifiers was confusion between the emergent vegetation and open water classes. A more thorough sampling procedure would eliminate some of these error but due to the spectral responses of these two land covers, the increase in accuracy might not be worth the time invested.

Flooding extent in the CRB exhibits variability from year to year. 2014 was the largest flood and matches with studies previously conduced in the CRB. 2015 experienced a small flood compared to past research in the area. 2016 rebounded but only marginally. The overall trend appears to be decreasing but 3 years of data is tough to ascertain a trend. Observing data collected from the region over the years by Burke et al. (2016), the CRB appears to experience

dry and wet cycles because 2005 experienced small flood as well. Whether this information from Burke et al. (2016) indicates the region experiences 10 year dry and wet cycles, more data needs to be collected to validate that claim.

Some limitations associated with the study include cloudy images in the study area that resulted in heavily misclassified imagery. Time-window 1 proved especially difficult to classify due to these clouds and rapidly changing tasseled cap component values for the land covers in the CRB. Furthermore, while areas within the study area accurately mapped the land covers, areas outside CRB were prone to high misclassification. The misclassification comes from burn scars and vegetation that exhibits similar tasseled cap component values similar to emergent vegetation.

Future work includes using a training library collected inter-annually so there is only one training library to classify imagery over many years. The purpose of one training library would be to determine if using training samples from multiple years results in comparative accuracy results to just using one training library per time-window. Using training samples across years would save more time in the collection of training samples and ultimately become a training library that can classify imagery for the MODIS time period stretching back to 2001. Expanding upon the one training library logic, another avenue of research relates to developing an integrated training library that could be applied to similar situations in other parts of the world. The integrated training library would have values collected from the CRB and then those values would be applied to classify imagery in another wetland that occurs in an arid region. Another area of research is using only one or two components of the Kauth-Thomas tasseled cap transformation to see if accuracy results are comparable to using all three components. The brightness and wetness components in particular map the training samples for the land covers in a linear fashion and could provide a more streamlined process of classification.

Bibliography

Adam, E., O. Mutanga, J. Odindi, and E. M. Abdel-Rahman. 2014. Land-use/cover Classification in a Heterogeneous Coastal Landscape using RapidEye Imagery: Evaluating the Performance of Random Forest and Support Vector Machines Classifiers. *International Journal of Remote Sensing* 35 (10): 3440-3458.

Amarnath, Giriraj and Ameer Rajah. 2015. An Evaluation of Flood Inundation Mapping from MODIS and ALOS Satellites for Pakistan. *Geomatics, Natural Hazards and Risk* 1-12.

Ben-Hur, A., and J. Weston. 2010. A User's Guide to Support Vector Machines. *Data Mining Techniques for the Life Sciences*: 223-239.

Bischof, Horst, Werner Schneider, and Axel J. Pinz. 1992. Multispectral Classification of Landsat-Images using Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing* 30 (3): 482-490.

Boschetti, Mirco, Francesco Nutini, Giacinto Manfron, Pietro Alessandro Brivio, and Andrew Nelson. 2014. Comparative Analysis of Normalized Difference Spectral Indices Derived from MODIS for Detecting Surface Water in Flooded Rice Cropping Systems. *PloS One* 9 (2): e88741.

Breiman, L. 2001. Random Forests. *Machine Learning* 45 (1): 5-32.

Burke, Jeri J., Narcisa G. Pricope, and James Blum. 2016. Thermal Imagery-Derived Surface Inundation Modeling to Assess Flood Risk in a Flood-Pulsed Savannah Watershed in Botswana and Namibia. *Remote Sensing* 8 (8): 676.

Chang, Chih-Chung and Chih-Jen Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2 (3): 27.

Chen, Yun, Chang Huang, Catherine Ticehurst, Linda Merrin, and Peter Thew. 2013. An Evaluation of MODIS Daily and 8-Day Composite Products for Floodplain and Wetland Inundation Mapping. *Wetlands* 33 (5): 823-835.

Chen, Yun, Bing Wang, Carmel A. Pollino, Susan M. Cuddy, Linda E. Merrin, and Chang Huang. 2014. Estimate of Flood Inundation and Retention on Wetlands using Remote Sensing and GIS. *Ecohydrology* 7 (5): 1412-1420.

Cortes, C., and V. Vapnik. 1995. Support-Vector Networks. *Machine Learning* 20 (3):273-297.

Dechka, JA, SE Franklin, MD Watmough, RP Bennett, and DW Ingstrup. 2002. Classification of Wetland Habitat and Vegetation Communities using Multi-Temporal Ikonos Imagery in Southern Saskatchewan. *Canadian Journal of Remote Sensing* 28 (5): 679-685.

Duro, Dennis C., Steven E. Franklin, and Monique G. Dubé. 2012. A Comparison of Pixel-Based and Object-Based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes using SPOT-5 HRG Imagery. *Remote Sensing of Environment* 118: 259-272.

Eisavi, Vahid, Saeid Homayouni, Ahmad Maleknezhad Yazdi, and Abbas Alimohammadi. 2015. Land Cover Mapping Based on Random Forest Classification of Multitemporal Spectral and Thermal Images. *Environmental Monitoring and Assessment* 187 (5): 291.

Funkenberg, Tim, Tran Thai Binh, Florian Moder, and Stefan Dech. 2014. The Ha Tien Plain–Wetland Monitoring using Remote-Sensing Techniques. *International Journal of Remote Sensing* 35 (8): 2893-2909.

Govaerts, Yves M., Michael M. Verstraete, Bernard Pinty, and Nadine Gobron. 1999. Designing Optimal Spectral Indices: A Feasibility and Proof of Concept Study. *International Journal of Remote Sensing* 20 (9): 1853-1873.

Halabisky, Meghan, L. Monika Moskal, Alan Gillespie, and Michael Hannam. 2016. Reconstructing Semi-Arid Wetland Surface Water Dynamics through Spectral Mixture Analysis of a Time Series of Landsat Satellite Images (1984–2011). *Remote Sensing of Environment* 177: 171-183.

Han, Xingxing, Xiaoling Chen, and Lian Feng. 2015. Four Decades of Winter Wetland Changes in Poyang Lake Based on Landsat Observations between 1973 and 2013. *Remote Sensing of Environment* 156: 426-437.

Hansen, Matthew, R. Dubayah, and R. DeFries. 1996. Classification Trees: An Alternative to Traditional Land Cover Classifiers. *International Journal of Remote Sensing* 17 (5): 1075-1081.

Huang, Chang, Yun Chen, and Jianping Wu. 2014. Mapping Spatio-Temporal Flood Inundation Dynamics at Large River Basin Scale using Time-Series Flow Data and MODIS Imagery. *International Journal of Applied Earth Observation and Geoinformation* 26: 350-362.

Ji, Lei, Li Zhang, and Bruce Wylie. 2009. Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *Photogrammetric Engineering & Remote Sensing* 75 (11): 1307-1317.

Liu, Jiantao, Quanlong Feng, Jianhua Gong, Jieping Zhou, and Yi Li. 2016. Land-Cover Classification of the Yellow River Delta Wetland Based on Multiple End-Member Spectral Mixture Analysis and a Random Forest Classifier. *International Journal of Remote Sensing* 37 (8): 1845-1867.

Kauth, Richard J. and GS Thomas. 1976. The Tasselled Cap--a Graphic Description of the Spectral-Temporal Development of Agricultural Crops as seen by Landsat.

Kavzoglu, T., and I. Colkesen. 2009. A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification. *International Journal of Applied Earth Observation and Geoinformation* 11 (5): 352-359.

Kumar, Lalit, Priyakant Sinha, and Subhashni Taylor. 2014. Improving Image Classification in a Complex Wetland Ecosystem through Image Fusion Techniques. *Journal of Applied Remote Sensing* 8 (1): 083616-083616.

Laborte, Alice G., Aileen A. Maunahan, and Robert J. Hijmans. 2010. Spectral Signature Generalization and Expansion Can Improve the Accuracy of Satellite Image Classification. *PloS One* 5 (5): e10516.

Li, Linlin, Anton Vrieling, Andrew Skidmore, Tiejun Wang, Antonio-Román Muñoz, and Eren Turak. 2015. Evaluation of MODIS Spectral Indices for Monitoring Hydrological Dynamics of a Small, Seasonally-Flooded Wetland in Southern Spain. *Wetlands* 35 (5): 851-864.

Li, Manqi, Jungho Im, and Colin Beier. 2013. Machine Learning Approaches for Forest Classification and Change Analysis using Multi-Temporal Landsat TM Images Over Huntington Wildlife Forest. *GIScience & Remote Sensing* 50 (4): 361-384.

Lobser, S. E., and W. B. Cohen. 2007. MODIS Tasselled Cap: Land Cover Characteristics Expressed through Transformed MODIS Data. *International Journal of Remote Sensing* 28 (22): 5079-5101.

Long, Stephanie, Temilola E. Fatoyinbo, and Frederick Policelli. 2014. Flood Extent Mapping for Namibia using Change Detection and Thresholding with SAR. *Environmental Research Letters* 9 (3): 035002.

Millard, Koreen, and Murray Richardson. 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing* 7 (7): 8489-8515.

Mitsch, W. J., and J. G. Gosselink. 1993. *Wetlands.* (2nd edn). New York: Van Nostrand Reinhold.

Mountrakis, G., J. Im, and C. Ogole. 2011. Support Vector Machines in Remote Sensing: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3): 247-259.

Ogilvie, Andrew, Gilles Belaud, Carole Delenne, Jean-Stéphane Bailly, Jean-Claude Bader, Aurélie Oleksiak, Luc Ferry, and Didier Martin. 2015. Decadal Monitoring of the Niger Inner Delta Flood Dynamics using MODIS Optical Data. *Journal of Hydrology* 523: 368-383.

Olofsson, Pontus, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. 2014. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sensing of Environment* 148: 42-57.

Otukei, John Richard, and Thomas Blaschke. 2010. Land Cover Change Assessment using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms. *International Journal of Applied Earth Observation and Geoinformation* 12: S27-S31.

Pal, Mahesh. and PM Mather. 2005. Support Vector Machines for Classification in Remote Sensing. *International Journal of Remote Sensing* 26 (5): 1007-1011.

Pal, Mahesh. 2008. Ensemble of Support Vector Machines for Land Cover Classification. *International Journal of Remote Sensing* 29 (10): 3043-3049.

Pal, Mahesh and Paul M. Mather. 2003. An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification. *Remote Sensing of Environment* 86 (4): 554-565.

Pricope, Narcisa G. 2013. Variable-Source Flood Pulsing in a Semi-Arid Transboundary Watershed: The Chobe River, Botswana and Namibia. *Environmental Monitoring and Assessmen*t 185 (2): 1883-1906.

Pricope, Narcisa G., Andrea E. Gaughan, John D. All, Michael W. Binford, and Lucas P. Rutina. 2015. Spatio-Temporal Analysis of Vegetation Dynamics in Relation to Shifting Inundation and Fire Regimes: Disentangling Environmental Variability from Land Management Decisions in a Southern African Transboundary Watershed. *Land* 4 (3): 627-655.

Pricope, Narcisa G. and Michael W. Binford. 2012. A Spatio-Temporal Analysis of Fire Recurrence and Extent for Semi-Arid Savanna Ecosystems in Southern Africa using Moderate-Resolution Satellite Imagery. *Journal of Environmental Management* 100: 72-85.

Schmid, T., M. Koch, J. Gumuzzio, and PM Mather. 2004. A Spectral Library for a Semi-Arid Wetland and its Application to Studies of Wetland Degradation using Hyperspectral and Multispectral Data. *International Journal of Remote Sensing* 25 (13): 2485-2496.

Shao, Yang and Ross S. Lunetta. 2012. Comparison of Support Vector Machine, Neural Network, and CART Algorithms for the Land-Cover Classification using Limited Training Data Points. *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 78-87.

Shaker, Ahmed, Wai Yeung Yan, and Nagwa El-Ashmawy. 2012. Panchromatic Satellite Image Classification for Flood Hazard Assessment. *Journal of Applied Research and Technology* 10 (6): 902-911.

Srivastava, Prashant K., Dawei Han, Miguel A. Rico-Ramirez, Michaela Bray, and Tanvir Islam. 2012. Selection of Classification Techniques for Land use/land Cover Change Investigation. *Advances in Space Research* 50 (9): 1250-1265.

Szantoi, Zoltan, Francisco Escobedo, Amr Abd-Elrahman, Scot Smith, and Leonard Pearlstine. 2013. Analyzing Fine-Scale Wetland Composition using High Resolution Imagery and Texture Features. *International Journal of Applied Earth Observation and Geoinformation* 23: 204-212.

Rokni, Komeil, Anaur Ahmad, Ali Selamat, and Sharifeh Hazini. 2014. Water Feature Extraction and Change Detection Using Mulititemporal Landsat Imagery. *Remote Sensing* 6: 4173-4189.

Wang, Jida, Yongwei Sheng, and Tak Shun D. Tong. 2014. Monitoring Decadal Lake Dynamics Across the Yangtze Basin Downstream of Three Gorges Dam. *Remote Sensing of Environment* 152: 251-269.

Wang, Lin, Iryna Dronova, Peng Gong, Wenbo Yang, Yingren Li, and Qing Liu. 2012. A New Time Series vegetation–water Index of phenological–hydrological Trait Across Species and Functional Types for Poyang Lake Wetland Ecosystem. *Remote Sensing of Environment* 125: 49-63.

Waske, B., and Sebastian van der Linden. 2008. Classifying Multilevel Imagery from SAR and Optical Sensors by Decision Fusion. *IEEE Transactions on Geoscience and Remote Sensing* 46 (5): 1457-1466.

Wright, Chris and Alisa Gallant. 2007. Improved Wetland Remote Sensing in Yellowstone National Park using Classification Trees to Combine TM Imagery and Ancillary Environmental Data. *Remote Sensing of Environment* 107 (4): 582-605.

Xu, Hanqiu. 2006. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *International Journal of Remote Sensing* 27 (14): 3025-3033.

Zhang, Caiyun and Zhixiao Xie. 2013. Object-Based Vegetation Mapping in the Kissimmee River Watershed using HyMap Data and Machine Learning Techniques. *Wetlands* 33 (2): 233-244.

Zomer, Robert J., Antonio Trabucco, and SL Ustin. 2009. Building Spectral Libraries for Wetlands Land Cover Classification and Hyperspectral Remote Sensing. *Journal of Environmental Management* 90 (7): 2170-2177.

Appendix A - Error Matrices for Time-Window Validation

Table A.1 March 14, 2014 decision tree error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	40	1	0	41	97.56%
Emergent	11	29	0	40	72.50%
Open Water	2	12	35	49	71.43%
Column Total	53	42	35	130	
Producer's Accuracy	75.47%	69.05%	100.00%		
Overall accuracy = (40 + 29 + 35)/130 = 0.80					
Kappa Accuracy = 0.70					

Table A.2 May 17, 2014 decision tree error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	36	0	0	36	100.00%
Emergent	3	45	2	50	90.00%
Open Water	0	10	40	50	80.00%
Column Total	39	55	42	136	
Producer's Accuracy	92.31%	81.82%	95.24%		
Overall accuracy = (36 + 45 + 40)/136 = 0.89					
Kappa Accuracy = 0.83					

Table A.3 June 18, 2014 decision tree error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	48	2	0	50	96.00%
Emergent	5	38	0	43	88.37%
Open Water	0	7	42	49	85.71%
Column Total	53	47	42	142	
Producer's Accuracy	90.57%	80.85%	100.00%		
Overall accuracy = (48 + 38 + 42)/142 = 0.90					
Kappa Accuracy = 0.85					

Table A.4 March 14, 2014 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	14	0	0	14	100.00%
Emergent	6	16	0	22	72.73%
Open Water	0	19	29	48	60.42%
Column Total	20	35	29	84	
Producer's Accuracy	70.00%	45.71%	100.00%		
Overall accuracy = (14 + 16 + 29)/84 = 0.70					
Kappa Accuracy = 0.54					

Table A.5 May 17, 2014 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	27	1	0	28	96.43%
Emergent	2	39	2	43	90.70%
Open Water	1	8	40	49	81.63%
Column Total	30	48	42	120	
Producer's Accuracy	90.00%	81.25%	95.24%		
Overall accuracy = (27 + 39 + 40)/120 = 0.88					
Kappa Accuracy = 0.82					

Table A.6 June 18, 2014 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	19	0	0	19	100.00%
Emergent	3	55	1	59	93.22%
Open Water	0	2	35	37	94.59%
Column Total	22	57	36	115	
Producer's Accuracy	86.36%	96.49%	97.22%		
Overall accuracy = (19 + 55 + 35)/115 = 0.95					
Kappa Accuracy = 0.91					

Table A.7 March 14, 2014 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	40	0	0	40	100.00%
Emergent	17	22	0	39	56.41%
Open Water	3	26	21	50	42.00%
Column Total	60	48	21	129	
Producer's Accuracy	66.67%	45.83%	100.00%		
Overall Accuracy = (40 + 22 + 21)/129 = 0.64					
Kappa Accuracy = 0.48					

Table A.8 May 17, 2014 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	35	0	0	35	100.00%
Emergent	5	42	1	48	87.50%
Open Water	0	11	39	50	78.00%
Column Total	40	53	40	133	
Producer's Accuracy	87.50%	79.25%	97.50%		
Overall Accuracy = (35 + 42 + 39)/133 = 0.87					
Kappa Accuracy = 0.81					

Table A.9 June 18, 2014 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	33	0	0	33	100.00%
Emergent	6	39	0	45	86.67%
Open Water	0	7	42	49	85.71%
Column Total	39	46	42	127	
Producer's Accuracy	84.62%	84.78%	100.00%		
Overall Accuracy = (33 + 39 + 42)/127 = 0.90					
Kappa Accuracy = 0.85					

Table A.10 March 14, 2015 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	60	1	0	61	98.36%
Emergent	9	28	0	37	75.68%
Open Water	2	3	45	50	90.00%
Column Total	71	32	45	148	
Producer's Accuracy	84.51%	87.50%	100.00%		
Overall Accuracy = (60 + 28 + 45)/148 = 0.90					
Kappa Accuracy = 0.84					

Table A.11 May 25, 2015 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	33	0	0	33	100.00%
Emergent	9	36	0	45	80.00%
Open Water	3	5	36	44	81.82%
Column Total	45	41	36	122	
Producer's Accuracy	73.33%	87.80%	100.00%		
Overall Accuracy = (33 + 36 + 36)/122 = 0.86					
Kappa Accuracy = 0.79					

Table A.12 June 26, 2015 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	45	0	0	45	100.00%
Emergent	4	37	1	42	88.10%
Open Water	2	0	41	43	95.35%
Column Total	51	37	42	130	
Producer's Accuracy	88.24%	100.00%	97.62%		
Overall Accuracy = (45 + 37 + 41)/130 = 0.95					
Kappa Accuracy = 0.92					

Table A.13 March 14, 2015 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	31	1	0	32	96.88%
Emergent	8	36	0	44	81.82%
Open Water	2	4	67	73	91.78%
Column Total	41	41	67	149	
Producer's Accuracy	75.61%	87.80%	100.00%		
Overall Accuracy = (31 + 36 + 67)/149 = 0.90					
Kappa Accuracy = 0.84					

Table A.14 May 25, 2015 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	34	1	0	35	97.14%
Emergent	5	56	0	61	91.80%
Open Water	1	2	40	43	93.02%
Column Total	40	59	40	139	
Producer's Accuracy	85.00%	94.92%	100.00%		
Overall Accuracy = (34 + 56 + 40)/139 = 0.94					
Kappa Accuracy = 0.90					

Table A.15 June 26, 2015 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	30	0	0	30	100.00%
Emergent	4	37	0	41	90.24%
Open Water	8	1	55	64	85.94%
Column Total	42	38	55	135	
Producer's Accuracy	71.43%	97.37%	100.00%		
Overall Accuracy = (30 + 37 + 55)/135 = 0.90					
Kappa Accuracy = 0.85					

Table A.16 March 14, 2015 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	45	0	1	46	97.83%
Emergent	17	28	0	45	62.22%
Open Water	0	3	47	50	94.00%
Column Total	62	31	48	141	
Producer's Accuracy	72.58%	90.32%	97.92%		
Overall Accuracy = (45 + 28 + 47)/141 = 0.85					
Kappa Accuracy = 0.78					

Table A.17 May 25, 2015 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	49	0	0	49	100.00%
Emergent	5	43	1	49	87.76%
Open Water	0	10	40	50	80.00%
Column Total	54	53	41	148	
Producer's Accuracy	90.74%	81.13%	97.56%		
Overall Accuracy = (49 + 43 + 40)/148 = 0.89					
Kappa Accuracy = 0.84					

Table A.18 June 26, 2015 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	49	2	1	52	94.23%
Emergent	7	38	2	47	80.85%
Open Water	6	4	32	42	76.19%
Column Total	62	44	35	141	
Producer's Accuracy	79.03%	86.36%	91.43%		
Overall Accuracy = (49 + 38 + 32)/141 = 0.84					
Kappa Accuracy = 0.76					

Table A.19 Feb 10, 2016 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	57	0	0	57	100.00%
Emergent	28	14	1	43	32.56%
Open Water	1	3	36	40	90.00%
Column Total	86	17	37	140	
Producer's Accuracy	66.28%	82.35%	97.30%		
Overall Accuracy = (57 + 14 + 36)/140 = 0.76					
Kappa Accuracy = 0.63					

Table A.20 May 24, 2016 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	36	0	0	36	100.00%
Emergent	4	38	1	43	88.37%
Open Water	0	8	42	50	84.00%
Column Total	40	46	43	129	
Producer's Accuracy	90.00%	82.61%	97.67%		
Overall Accuracy = (36 + 38 + 42)/129 = 0.90					
Kappa Accuracy = 0.85					

Table A.21 June 25, 2016 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	37	0	0	37	100.00%
Emergent	4	44	1	49	89.80%
Open Water	1	6	47	54	87.04%
Column Total	42	50	48	140	
Producer's Accuracy	88.10%	88.00%	97.92%		
Overall Accuracy = (37 + 44 + 47)/140 = 0.91					
Kappa Accuracy = 0.87					

Table A.22 February 10, 2016 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	41	0	0	41	100.00%
Emergent	20	23	0	43	53.49%
Open Water	0	1	57	58	98.28%
Column Total	61	24	57	142	
Producer's Accuracy	67.21%	95.83%	100.00%		
Overall Accuracy = (41 + 23 + 57)/142 = 0.85					
Kappa Accuracy = 0.78					

Table A.23 May 24, 2016 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	36	0	0	36	100.00%
Emergent	4	52	0	56	92.86%
Open Water	2	3	38	43	88.37%
Column Total	42	55	38	135	
Producer's Accuracy	85.71%	94.55%	100.00%		
Overall Accuracy = (36 + 52 + 38)/135 = 0.93					
Kappa Accuracy = 0.90					

Table A.24 June 25, 2016 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	50	3	0	53	94.34%
Emergent	5	39	2	46	84.78%
Open Water	9	1	42	52	80.77%
Column Total	64	43	44	151	
Producer's Accuracy	78.13%	90.70%	95.45%		
Overall Accuracy = (50 + 39 + 42)/151 = 0.87					
Kappa Accuracy = 0.80					

Table A.25 Feb 10, 2016 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	54	0	0	54	100.00%
Emergent	25	13	0	38	34.21%
Open Water	3	7	41	51	80.39%
Column Total	82	20	41	143	
Producer's Accuracy	65.85%	65.00%	100.00%		
Overall Accuracy = (54 + 13 + 41)/143 = 0.76					
Kappa Accuracy = 0.62					

Table A.26 May 24, 2016 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	40	1	3	44	90.91%
Emergent	3	46	0	49	93.88%
Open Water	1	10	37	48	77.08%
Column Total	44	57	40	141	
Producer's Accuracy	90.91%	80.70%	92.50%		
Overall Accuracy = (40 + 46 + 37)/141 = 0.87					
Kappa Accuracy = 0.81					

Table A.27 June 25, 2016 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	48	2	1	51	94.12%
Emergent	7	49	0	56	87.50%
Open Water	3	2	44	49	89.80%
Column Total	58	53	45	156	
Producer's Accuracy	82.76%	92.45%	97.78%		
Overall Accuracy = (48 + 49 + 44)/156 = 0.90					
Kappa Accuracy = 0.86					

Appendix B - Error Matrices for Time-Window Discussion

Table B.1 March 14, 2015 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	59	0	0	59	100.00%
Emergent	11	52	2	65	80.00%
Open Water	3	10	60	73	82.19%
Column Total	73	62	62	197	
Producer's Accuracy	80.82%	83.87%	96.77%		
Overall Accuracy = (59 + 52 + 60)/197 = 0.87					
Kappa Accuracy = 0.80					

Table B.2 March 14, 2015 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	66	0	0	66	100.00%
Emergent	14	43	2	59	72.88%
Open Water	4	6	65	75	86.67%
Column Total	84	49	67	200	
Producer's Accuracy	78.57%	87.76%	97.01%		
Overall Accuracy = (66 + 43 + 65)/200 = 0.87					
Kappa Accuracy = 0.80					

Table B.3 March 14, 2015 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	63	0	0	63	100.00%
Emergent	12	40	0	52	76.92%
Open Water	2	9	64	75	85.33%
Column Total	77	49	64	190	
Producer's Accuracy	81.82%	81.63%	100.00%		
Overall Accuracy = (63 + 40 + 64)/190 = 0.88					
Kappa Accuracy = 0.82					

Table B.4 May 25, 2015 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	65	0	0	65	100.00%
Emergent	8	59	2	69	85.51%
Open Water	0	15	58	73	79.45%
Column Total	73	74	60	207	
Producer's Accuracy	89.04%	79.73%	96.67%		
Overall Accuracy = (65 + 59 + 58)/207 = 0.88					
Kappa Accuracy = 0.82					

Table B.5 May 25, 2015 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	56	0	0	56	100.00%
Emergent	24	26	0	50	52.00%
Open Water	3	16	55	74	74.32%
Column Total	83	42	55	180	
Producer's Accuracy	67.47%	61.90%	100.00%		
Overall Accuracy = (56 + 26 + 55)/180 = 0.76					
Kappa Accuracy = 0.64					

Table B.6 May 25, 2015 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	65	0	0	65	100.00%
Emergent	12	49	0	61	80.33%
Open Water	4	8	52	64	81.25%
Column Total	81	57	52	190	
Producer's Accuracy	80.25%	85.96%	100.00%		
Overall Accuracy = (65 + 49 + 52)/190 = 0.87					
Kappa Accuracy = 0.81					

Table B.7 June 26, 2015 ML error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	60	1	1	62	96.77%
Emergent	7	57	0	64	89.06%
Open Water	9	6	48	63	76.19%
Column Total	76	64	49	189	
Producer's Accuracy	78.95%	89.06%	97.96%		
Overall Accuracy = (60 + 57 + 48)/189 = 0.87					
Kappa Accuracy = 0.81					

Table B.8 June 26, 2015 DT error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	64	2	0	66	96.97%
Emergent	10	49	1	60	81.67%
Open Water	3	5	52	60	86.67%
Column Total	77	56	53	186	
Producer's Accuracy	83.12%	87.50%	98.11%		
Overall Accuracy = (64 + 49 + 52)/186 = 0.89					
Kappa Accuracy = 0.83					

Table B.9 June 26, 2015 SVM error matrix

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	66	0	0	66	100.00%
Emergent	11	47	0	58	81.03%
Open Water	5	2	56	63	88.89%
Column Total	82	49	56	187	
Producer's Accuracy	80.49%	95.92%	100.00%		
Overall Accuracy = (66 + 47 + 56)/187 = 0.90					
Kappa Accuracy = 0.85					

Table B.10 March 14, 2015 ML error matrix using time-window 2 training samples

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	56	7	0	63	88.89%
Emergent	8	52	5	65	80.00%
Open Water	0	20	41	61	67.21%
Column Total	64	79	46	189	
Producer's Accuracy	87.50%	65.82%	89.13%		
Overall Accuracy = (56 + 52 + 41)/189 = 0.79					
Kappa Accuracy = 0.68					

Table B.11 March 14, 2015 DT error matrix using time-window 2 training samples

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	60	2	0	62	96.77%
Emergent	17	40	2	59	67.80%
Open Water	0	20	55	75	73.33%
Column Total	77	62	57	196	
Producer's Accuracy	77.92%	64.52%	96.49%		
Overall Accuracy = (60 + 40 + 55)/196 = 0.79					
Kappa Accuracy = 0.69					

Table B.12 March 14, 2015 SVM error matrix using time-window 2 training samples

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	63	4	0	67	94.03%
Emergent	10	48	5	63	76.19%
Open Water	3	18	54	75	72.00%
Column Total	76	70	59	205	
Producer's Accuracy	82.89%	68.57%	91.53%		
Overall Accuracy = (63 + 48 + 54)/205 = 0.80					
Kappa Accuracy = 0.71					

Table B.13 June 26, 2015 ML error matrix using time-window 2 training samples

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	60	6	0	66	90.91%
Emergent	11	54	4	69	78.26%
Open Water	2	11	57	70	81.43%
Column Total	73	71	61	205	
Producer's Accuracy	82.19%	76.06%	93.44%		
Overall Accuracy = (60 + 54 + 57)/205 =0.83					
Kappa Accuracy = 0.75					

Table B.14 June 26, 2015 DT error matrix using time-window 2 training samples

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	62	5	0	67	92.54%
Emergent	17	40	3	60	66.67%
Open Water	1	9	56	66	84.85%
Column Total	80	54	59	193	
Producer's Accuracy	77.50%	74.07%	94.92%		
Overall Accuracy = (62 + 40 + 56)/193 =0.82					
Kappa Accuracy = 0.73					

Table B.15 June 26, 2015 SVM error matrix using time-window 2 training samples

	Land	Emergent	Open Water	Row Total	User's Accuracy
Land	60	5	0	65	92.31%
Emergent	11	52	5	68	76.47%
Open Water	5	8	49	62	79.03%
Column Total	76	65	54	195	
Producer's Accuracy	78.95%	80.00%	90.74%		
Overall Accuracy = (60 + 52 + 49)/189 = 0.83					
Kappa Accuracy = 0.74					

Appendix C - Decision Trees

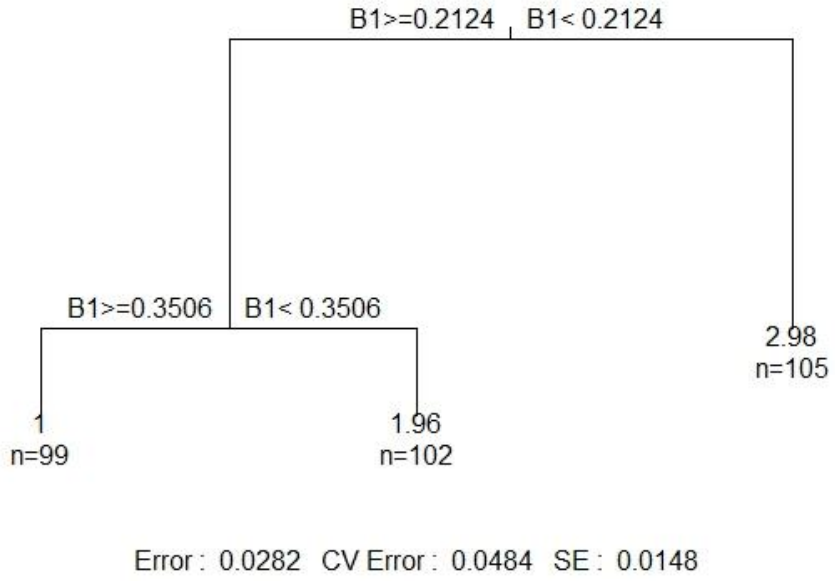


Figure C.1 DT for time-window 1 2014

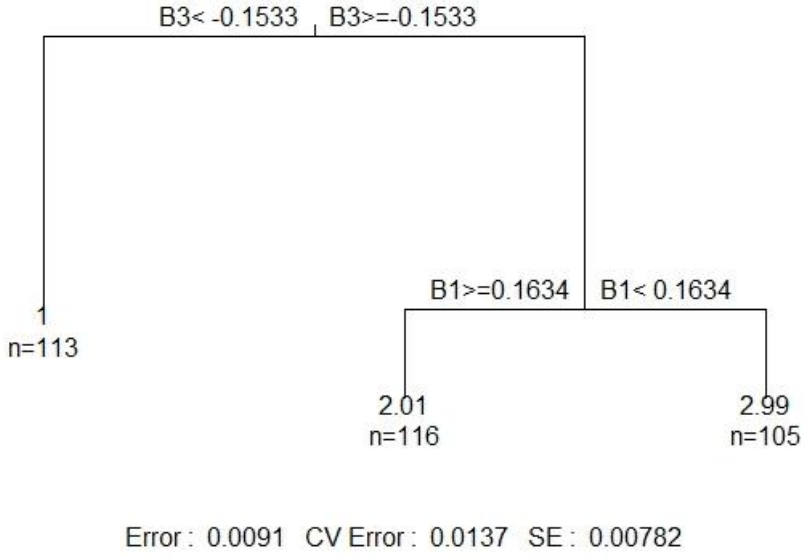


Figure C.2 DT for time-window 2 2014

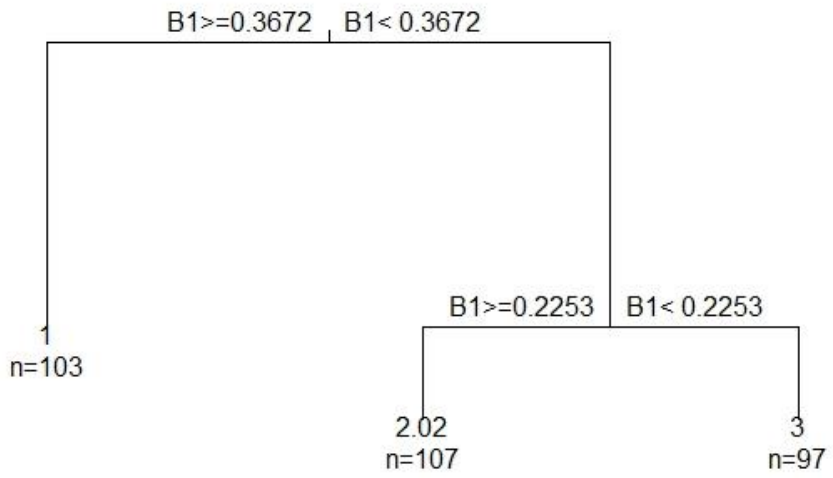


Figure C.3 DT for time-window 3 2014

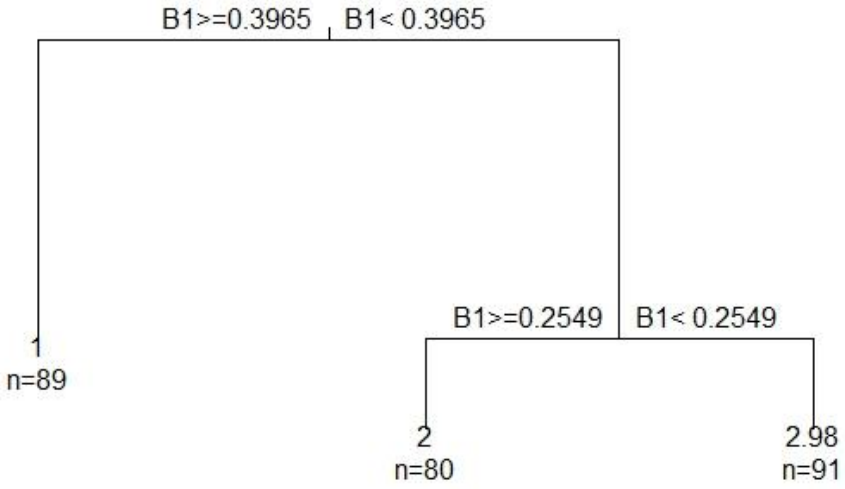


Figure C.4 DT for time-window 1 2015

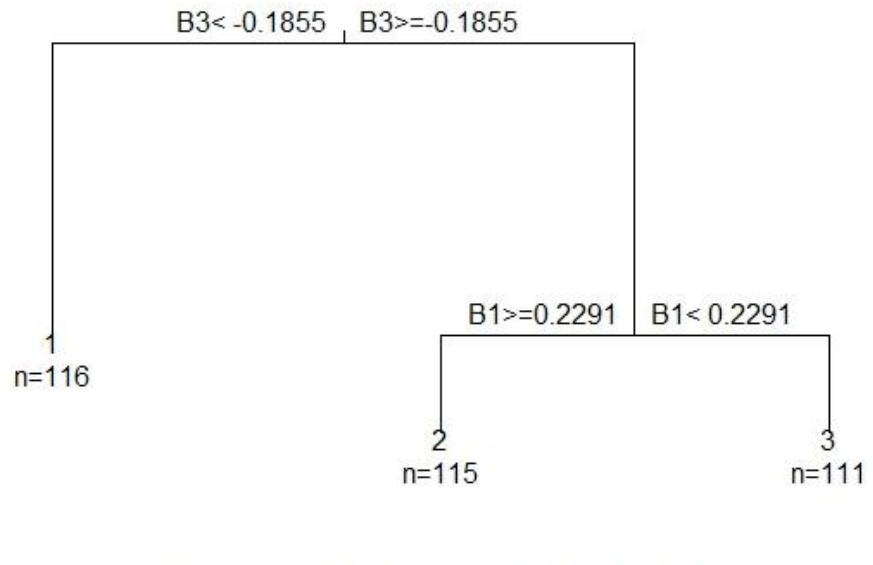


Figure C.5 DT for time-window 2 2015

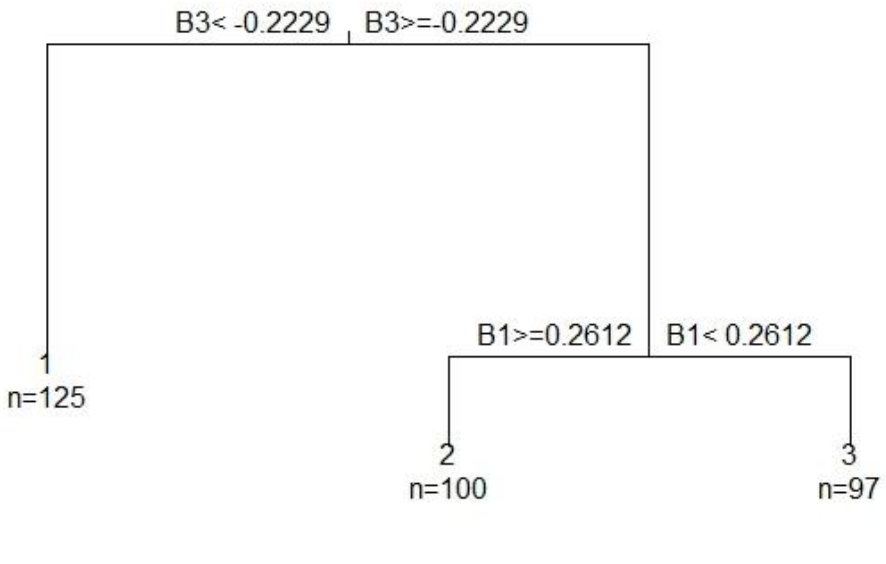


Figure C.6 DT for time-window 3 2015

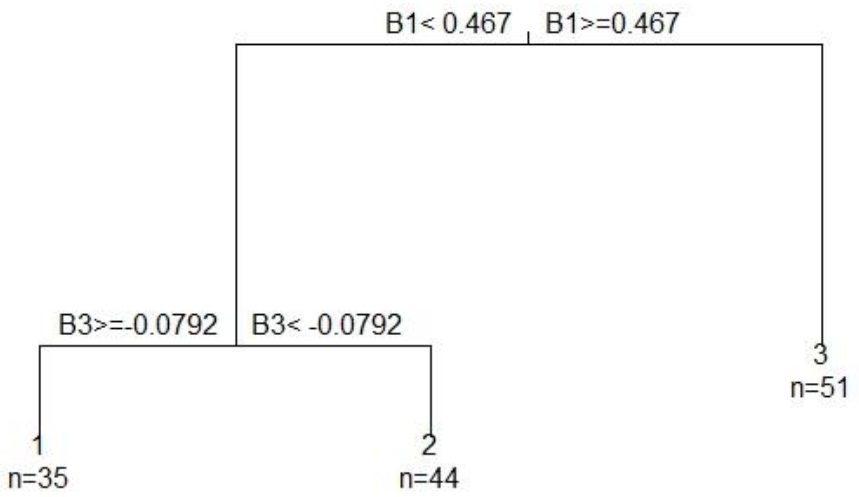


Figure C.7 DT for time-window 1 2016

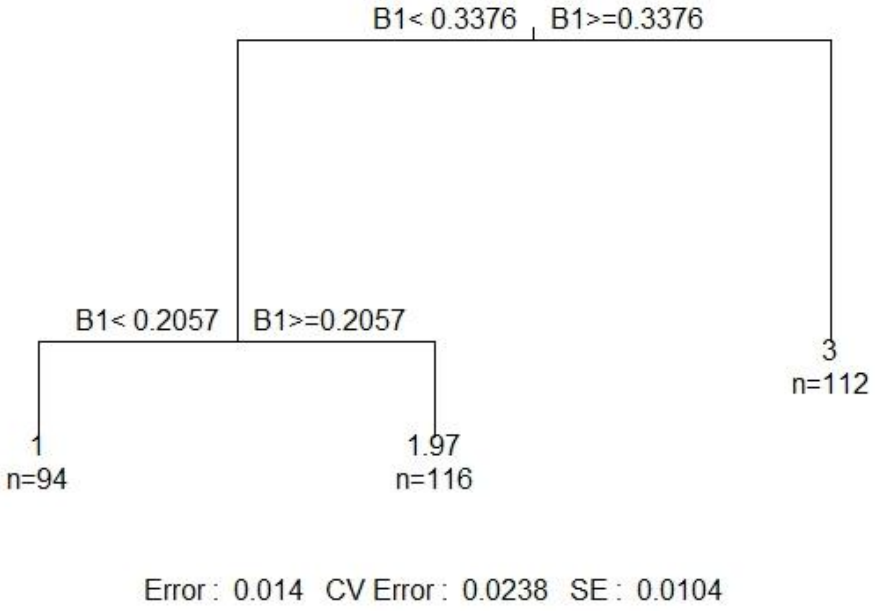


Figure C.8 DT for time-window 2 2016

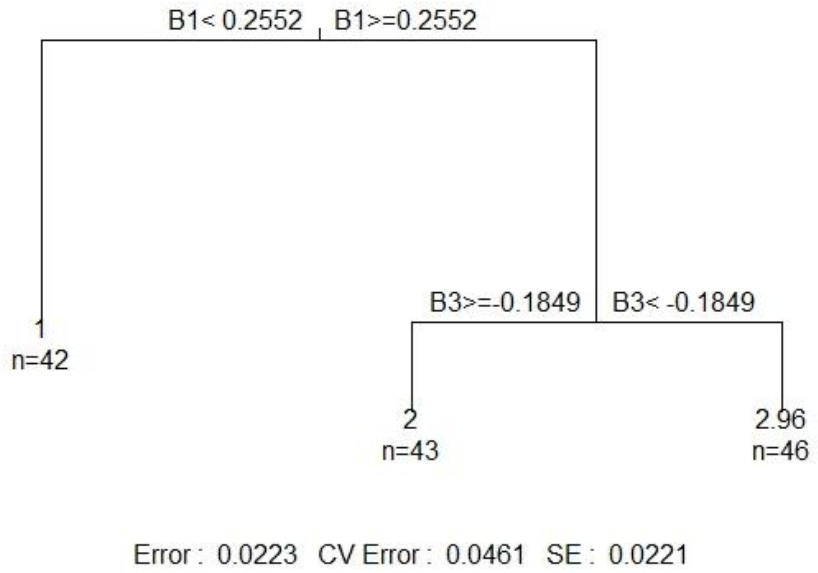


Figure C.9 DT for time-window 3 2016

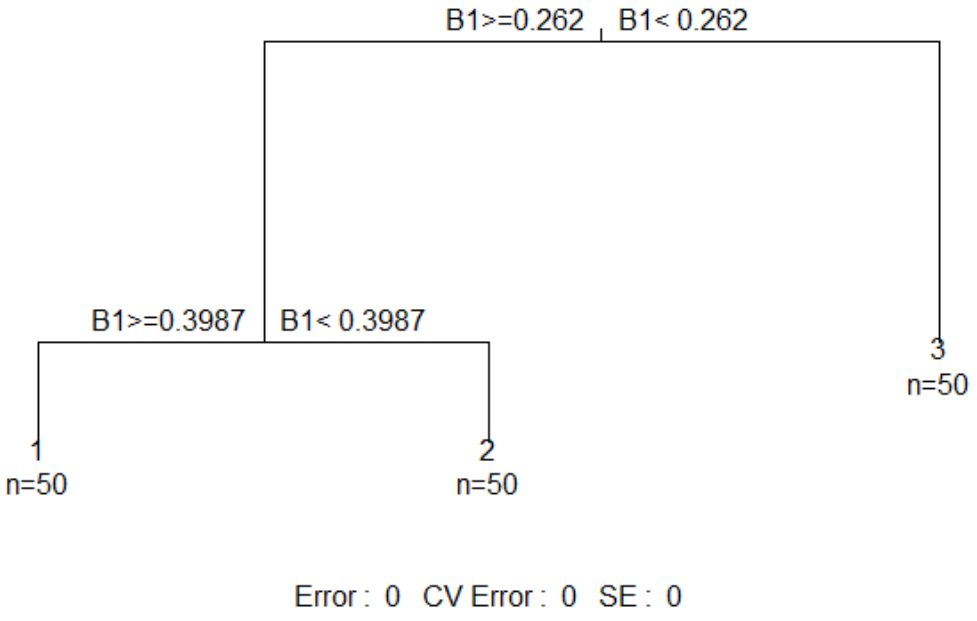


Figure C.10 DT for time-window 1 2015 when comparing traditional classification versus time-window method

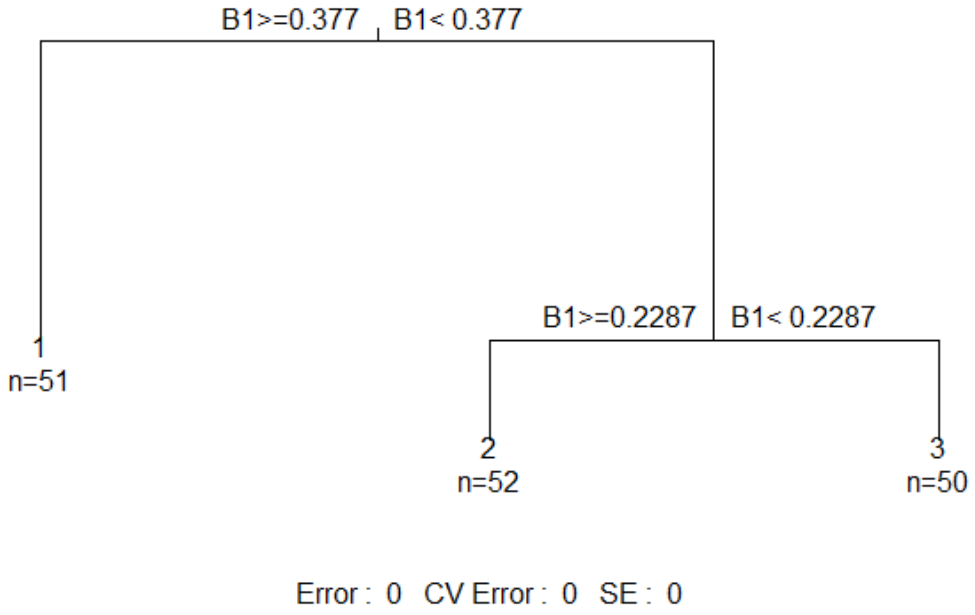


Figure C.11 DT for time-window 2 2015 when comparing traditional classification versus time-window method

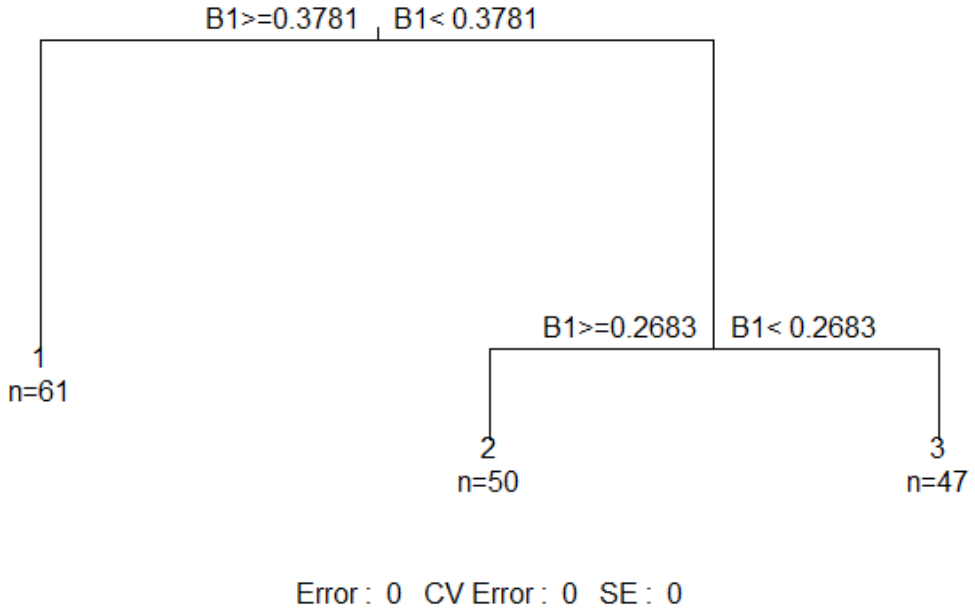


Figure C.12 DT for time-window 1 2015 when comparing traditional classification versus time-window method

Appendix D - Code for SVMs

Code run to parse the tasseled cap component values of each land for a given time-window

```
; These training samples need to be in the libsvm format
training_sample =
"F:\Chobe_Project\MOD09A1\Discussion_Classification\TrainingSamples_TimeWindow3_SVM.t
xt"
OPENR, lun, training_sample, /GET_LUN
; Read one line at a time, saving the result into array
line = ''
txt_array = []
WHILE NOT EOF(lun) DO BEGIN & $
  READF, lun, line & $
  ;array = [array, line] & $
  a = STRSPLIT(line, /EXTRACT)
; Grab the brightness, greenness, and wetness values from the text file and exclude
the 1:, 2:, 3:, numbers
this_line_txt = [float(a[0]), float(a[2]), float(a[4]), float(a[6])]
txt_array = [[txt_array], [this_line_txt]]
ENDWHILE
; Close the file and free the file unit
FREE_LUN, lun


End
```

Code run to generate a classified SVM image, using the tasseled cap component values for a given time-window, as well as incorporating the SVMs gamma and penalty parameters.

```
; Create classification image using a spectral library
; Coded by Jida Wang and Mitch Braget on 23 November, 2016
;
================================================================================
===============

;;=======comment out if tested beyond the first image===============
;; Location of ROI file that includes water, emergent vegetation, and non-water
training samples from the extended images
ROI_file = 'F:\Chobe_Project\TimeWindow3_2015_SVM.roi'
;
;; Applying ROI for training with the ROIs you create on the extended images
;ENVI_RESTORE_ROIS, ROI_file
;roi_ids = ENVI_GET_ROI_IDS()  ;SOMETHING REDUNDANT ADDED IN HERE
;roi_dims = ENVI_GET_ROI_DIMS_PTR(roi_ids[0])
;ENVI_GET_ROI_INFORMATION, roi_ids ,NPTS=NPTS_value
;
POS = [0,1,2]
;;num_classes = 3
```

```
;;mean_value = FLTARR(N_ELEMENTS(pos), num_classes)
;;cov_value = FLTARR(N_ELEMENTS(pos),N_ELEMENTS(pos),num_classes)
;;FOR j=0, num_classes-1 DO BEGIN
;;  roi_dims2=[ENVI_GET_ROI_DIMS_PTR(roi_ids[j]),0,0,0,0]
;;  ENVI_DOIT, 'envi_stats_doit', fid=modified_image_FID, pos=pos, $
;;    dims=roi_dims2, comp_flag=4, mean=c_mean, $
;;    stdv=c_stdv, cov=c_cov
;;  MEAN_value[0,j] = c_mean
;;  cov_value[0,0,j] = c_cov
;;ENDFOR
;;================================================================================
======
```

```
;Mask Shapefile: full path to the shapefile of your customize waterbody region
Mask_Shapefile = 'F:\Chobe_Project\ValidationData\CRB_Mask.shp'
;1. Read txt file of spectral library of a certain time-window's value into an array
; These training samples need to be in the libsvm format
training_sample =
"F:\Chobe_Project\MOD09A1\Discussion_Classification\TrainingSamples_TimeWindow3_SVM.t
xt"
OPENR, lun, training_sample, /GET_LUN
; Read one line at a time, saving the result into array
line = ''
txt_array = []
WHILE NOT EOF(lun) DO BEGIN & $
  READF, lun, line & $
  ;array = [array, line] & $
  a = STRSPLIT(line, /EXTRACT)
; Grab the brightness, greenness, and wetness values from the text file and exclude
the 1:, 2:, 3:, numbers
this_line_txt = [float(a[0]), float(a[2]), float(a[4]), float(a[6])]
txt_array = [[txt_array], [this_line_txt]]
ENDWHILE
; Close the file and free the file unit
FREE_LUN, lun

;2. open any MODIS (Brightness, greenness, and wetness) image
Image_File =
'F:\Chobe_Project\MOD09A1\2015_Images\June_26_2015\Kauth_Thomas_June262015.hdr'
;Read in the image
ENVI_OPEN_FILE, Image_File,/NO_INTERACTIVE_QUERY,/NO_REALIZE, R_FID = IMAGE_FID
;Read image basic properties
ENVI_FILE_QUERY, IMAGE_FID, NB=NB, DIMS=DIMS, NS = NS, NL=NL, BNAMES = BNAMES,
XSTART= Image_Xstart, YSTART = Image_Ystart
MapInfo= ENVI_GET_MAP_INFO(FID = IMAGE_FID) ;Image projection
ImageProjection = ENVI_GET_PROJECTION(FID = IMAGE_FID, PIXEL_SIZE = ImagePixSize,
UNITS = ProjectUnits)
; Grab each of the KT bands
brightness = Float(ENVI_GET_DATA( FID=IMAGE_FID, DIMS=DIMS, POS=[0]))
greenness = Float(ENVI_GET_DATA( FID=IMAGE_FID, DIMS=DIMS, POS=[1]))
```

```
wetness = Float(ENVI_GET_DATA( FID=IMAGE_FID, DIMS=DIMS, POS=[2]))


;append the three bands by training sample to the brightness band of the image,
extending it my each land cover; water, emergent vegetation, non-water
; BRIGHTNESS
brightness_training = txt_array[1,*]
;land cover 1
cover_1_indices = where(txt_array[0,*] EQ 1)
brightness_cover1_training = brightness_training[cover_1_indices]
;land cover 2
cover_2_indices = where(txt_array[0,*] EQ 2)
brightness_cover2_training = brightness_training[cover_2_indices]
;land cover 3
cover_3_indices = where(txt_array[0,*] EQ 3)
brightness_cover3_training = brightness_training[cover_3_indices]
filled_nulls = fltarr(NS-n_elements(brightness_cover1_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover1 = [brightness_cover1_training, filled_nulls]
filled_nulls = fltarr(NS-n_elements(brightness_cover2_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover2 = [brightness_cover2_training, filled_nulls]
filled_nulls = fltarr(NS-n_elements(brightness_cover3_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover3 = [brightness_cover3_training, filled_nulls]
appended_covers = [[appended_cover1], [appended_cover2], [appended_cover3]]
; create the extended brightness band with the land covers
brightness_modified = [[brightness],[appended_covers]]

; GREENNESS append the three bands by training sample to the greenness band of the
image,
; extending it my each land cover; water, emergent vegetation, non-water
greenness_training = txt_array[2,*]
;land cover 1
;cover_1_indices = where(txt_array[0,*] EQ 1)
greenness_cover1_training = greenness_training[cover_1_indices]
;land cover 2
;cover_2_indices = where(txt_array[0,*] EQ 2)
greenness_cover2_training = greenness_training[cover_2_indices]
;land cover 3
;cover_3_indices = where(txt_array[0,*] EQ 3)
greenness_cover3_training = greenness_training[cover_3_indices]
filled_nulls = fltarr(NS-n_elements(greenness_cover1_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover1 = [greenness_cover1_training, filled_nulls]
filled_nulls = fltarr(NS-n_elements(greenness_cover2_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover2 = [greenness_cover2_training, filled_nulls]
filled_nulls = fltarr(NS-n_elements(greenness_cover3_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover3 = [greenness_cover3_training, filled_nulls]
appended_covers = [[appended_cover1], [appended_cover2], [appended_cover3]]
; create the extended greenness band with the three land covers
greenness_modified = [[greenness],[appended_covers]]
```

```
; WETNESS append the three bands by training sample to the wetness band of the image,
; extending it my each land cover; water, emergent vegetation, non-water
wetness_training = txt_array[3,*]
;land cover 1
;cover_1_indices = where(txt_array[0,*] EQ 1)
wetness_cover1_training = wetness_training[cover_1_indices]
;land cover 2
;cover_2_indices = where(txt_array[0,*] EQ 2)
wetness_cover2_training = wetness_training[cover_2_indices]
;land cover 3
;cover_3_indices = where(txt_array[0,*] EQ 3)
wetness_cover3_training = wetness_training[cover_3_indices]
filled_nulls = fltarr(NS-n_elements(wetness_cover1_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover1 = [wetness_cover1_training, filled_nulls]
filled_nulls = fltarr(NS-n_elements(wetness_cover2_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover2 = [wetness_cover2_training, filled_nulls]
filled_nulls = fltarr(NS-n_elements(wetness_cover3_training),1)
filled_nulls[0:n_elements(filled_nulls)-1] = !Values.F_NaN
appended_cover3 = [wetness_cover3_training, filled_nulls]
appended_covers = [[appended_cover1], [appended_cover2], [appended_cover3]]
; create the extended wetness band with the three land covers
wetness_modified = [[wetness],[appended_covers]]

; stack the three extended KT band together, the double [[]] signified a new band
KT_composite = [[[brightness_modified]], [[greenness_modified]],
[[wetness_modified]]]
new_dimension = size(KT_composite,/dimension)
envi_write_envi_file, KT_composite, ns=new_dimension[0], nl=new_dimension[1], R_FID =
modified_image_FID, $
  data_type = 4, file_type=envi_file_type('ENVI standard'), bnames = 'brightness
greeness wetness', $
  offset = 0, data_ignore_value= !Values.F_NaN, in_memory=1;,
out_name=output_dat_name, map_info=map_info,  /no_open
ENVI_FILE_QUERY, modified_image_FID, NB= modified_NB, DIMS= modified_DIMS, NS =
modified_NS, NL= modified_NL, BNAMES = modified_BNAMES;, XSTART= Image_Xstart, YSTART
= Image_Ystart

; extract the mask region
Mask_image = Mask_from_Shapefile(IMAGE_FID, NS, NL, Mask_Shapefile)
; Apply mask so only a small region is classified with SVM
KT_composite[where(Mask_image EQ 0)] = !Values.F_NaN

BAND_NAME = ['class']
;ENVI_DOIT, 'ENVI_SVM_DOIT', DIMS=modified_DIMS, FID=modified_image_FID, M_FID =
MASK_FID, M_POS = 0, /IN_MEMORY, OUT_BNAME=BAND_NAME, POS=POS, THRESH = 0.91, $
;  R_FID=class_lake_FID, ROI_IDS=roi_ids ;[, RULE_FID=variable] [, /RULE_IN_MEMORY]
[, RULE_OUT_NAME=string] [, THRESH=value]

; Run SVM with the radial basis function kernel, with the parameters gained from
libsvm
ENVI_DOIT, 'ENVI_SVM_DOIT', DIMS=modified_DIMS, FID=modified_image_FID, /IN_MEMORY,
OUT_BNAME=BAND_NAME, POS=POS, THRESH = 0.0, $
```

```
  R_FID=class_FID, ROI_IDS=roi_ids, KERNEL_TYPE = 2, KERNEL_GAMMA = 2, PENALTY = 0.5
;[, RULE_FID=variable] [, /RULE_IN_MEMORY] [, RULE_OUT_NAME=string] [, THRESH=value]

; truncate off the three extra lines and get the classified land covers back to the
0,1,2, positions for correct labeling
class_result = long(ENVI_GET_DATA(FID=class_FID, DIMS=modified_DIMS, POS=[0]))
class_result = class_result[*, 0:NL-1]
class_result = class_result-1
;new_class_result[where(class_result eq 1)] = 0
;new_class_result[where(class_result eq 2)] = 1
;new_class_result[where(class_result eq 3)] = 2

; Save the class image into drive
output =
'F:\Chobe_Project\MOD09A1\Discussion_Classification\SVM_26June2015_scaled.dat'
CLASS_NAMES     = ['Non-water','Emergent vegetation', 'Water'] ; Name of the 3
classes of land cover
DESCRIP         = 'MODIS-KT derived'
LOOKUP          = [[255,0,0],[0,255,0],[0,0,255]] ; Red, green, blue colors
BAND_NAME       = ['mask']
FILE_TYPE       = ENVI_FILE_TYPE('ENVI Classification')
IN_MEMORY       = LONARR(1)
; write the classified image to file
ENVI_WRITE_ENVI_FILE, class_result, BNAMES=['Classified Result'], NUM_CLASSES=3,
CLASS_NAMES=CLASS_NAMES, LOOKUP=LOOKUP,$
  DESCRIP=DESCRIP, FILE_TYPE=FILE_TYPE, MAP_INFO=MapInfo, PIXEL_SIZE=ImagePixSize,$
  UNITS=ProjectUnits, XSTART=Image_xstart, YSTART=Image_ystart, R_FID=
FINAL_CLASS_FID, out_name = output
end
```